Amino acids

Bischoff, Rainer; Schlueter, Hartmut

Published in:
Journal of proteomics

DOI:
10.1016/j.jprot.2012.01.041

Publication date:
2012

JOURNAL OF PROTEOMICS 75 (2012) 2275–2296

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

Tutorial

Amino acids: Chemistry, functionality and selected non-enzymatic post-translational modifications☆

Rainer Bischoff[a,*], *Hartmut Schlüter*[b]

[a]*University of Groningen, Department of Pharmacy, Analytical Biochemistry, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands*
[b]*Institute of Clinical Chemistry, University Medicine Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany*

ARTICLE INFO

Article history:
Received 12 October 2011
Accepted 31 January 2012
Available online 22 February 2012

Keywords:
Amino acid chemistry
Post-translational modifications
Protein species
Thiol group
Deamidation
Carbonylation

ABSTRACT

The ultimate goal of proteomics is determination of the exact chemical composition of protein species, including their complete amino acid sequence and the identification of each modified side chain, in every protein in a biological sample and their quantification. We are still far from achieving this goal due to limitations in analytical methodology and data analysis but also due to the fact that we surely have not discovered all amino acid modifications that occur in nature. To detect modified side chains and to discover new, still unknown amino acid derivatives, an understanding of the chemistry of the reactive groups of amino acids is mandatory. This tutorial focuses on the chemistry of the amino acid side chains and addresses non-enzymatic modifications. By highlighting some exemplary reactions a glimpse of the huge diversity of modified amino acids provides the reader with sufficient insight into amino acid chemistry to raise the awareness for unexpected side chain modifications. We further introduce the reader to a terminology, which enables the comprehensive description of the exact chemical composition of a protein species, including its full amino acid sequence and all modifications of its amino acid side chains. This Tutorial is part of the International Proteomics Tutorial Programme (IPTP number 10).

1. Introduction

1.1. *Historical background*

Amino acids are the primary building blocks of proteins. Of the over 300 naturally occurring amino acids, 22 constitute the monomer units of proteins [1], which are chemically linked via peptide bonds in genetically predefined sequences to constitute the backbone of all known proteins. Their properties range from acidic to basic and from hydrophilic to hydrophobic depending on the respective side chain R (Figs. 1–9). In 1806 Vauquelin and Robiquet isolated the first amino acid as a substance from *Asparagus sativus*, which they named asparagine [2]. Some years later Braconnot discovered glycine from an acid hydrolysate of gelatin [3]. Threonine was described by Rose et al. as an essential food ingredient in 1935 [4] and selenocysteine and pyrrolysine, the last two of the 22 proteinogenic amino acids to be discovered, were reported in 1986 [5] and 2002, respectively [6].

☆ This Tutorial is part of the International Proteomics Tutorial Programme (IPTP number 10). Details can be found at: http://www.proteomicstutorials.org/.

* *Corresponding author*. Tel.: +31 50 3633338.
E-mail address: r.p.h.bischoff@rug.nl (R. Bischoff).

1874-3919
doi:10.1016/j.jprot.2012.01.041

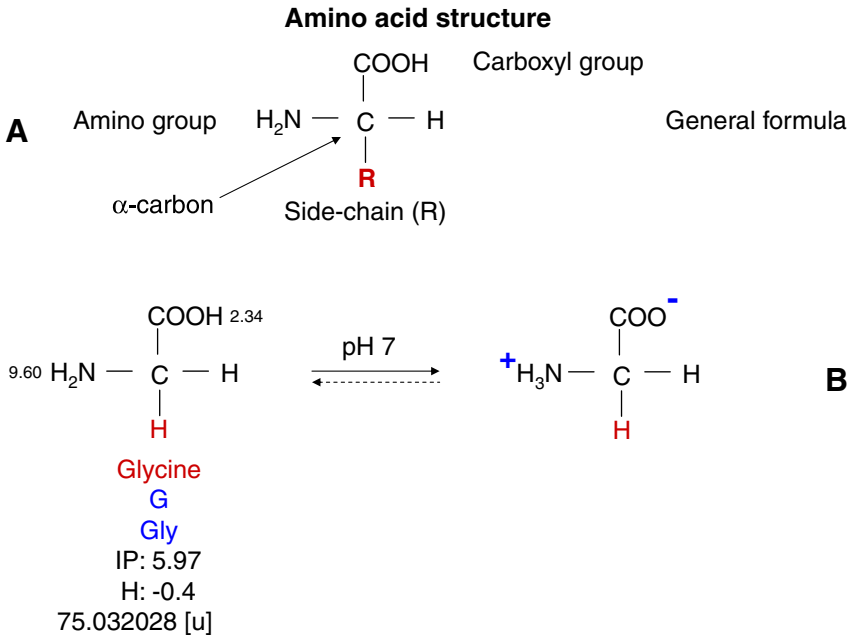

Fig. 1 – (A) Chemical structure of the amino acid glycine with the side chain R = H. (B) Amino acids generally occur as zwitterions at physiological pH due to partial ionization of the carboxylic acid and the amino group. The formulas present the ionization state predominating at pH 7. The hydropathy index (H) can be used to predict the tendency of an amino acid to seek a hydrophobic (values >1) or an aqueous (values <1) environment. Ionization state, isoelectric points (IP), dissociation constant (pK_a) values and hydropathy index from "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

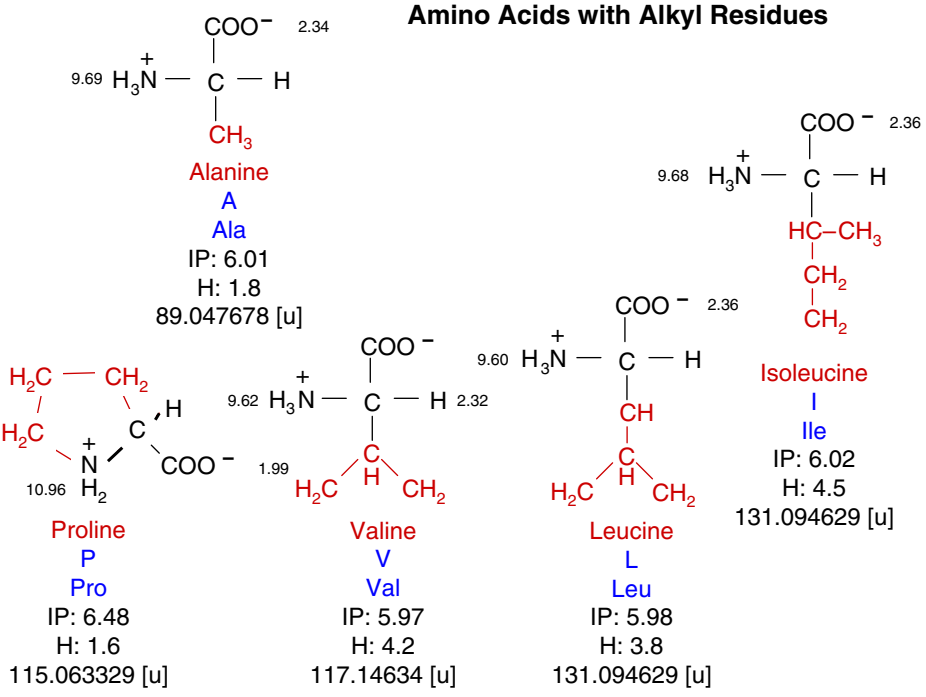

Fig. 2 – Proteinogenic amino acids with alkyl residues. The formulas present the ionization state predominating at pH 7. The hydropathy index (H) can be used to predict the tendency of an amino acid to seek a hydrophobic (values >1) or an aqueous (values <1) environment. Ionization state, IP, pK values and hydropathy index from "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

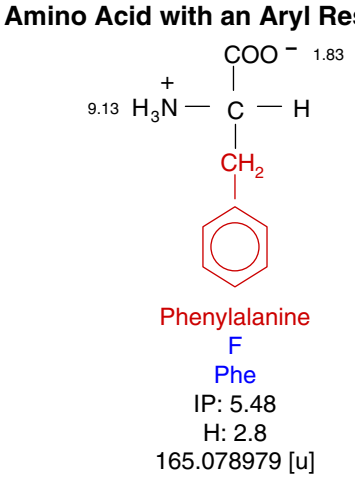

Fig. 3 – Proteinogenic amino acid with an aryl residue. The formulas present the ionization state predominating at pH 7. The hydropathy index (H) can be used to predict the tendency of an amino acid to seek a hydrophobic (values >1) or an aqueous (values <1) environment. Ionization state, IP, pK values and hydropathy index from "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

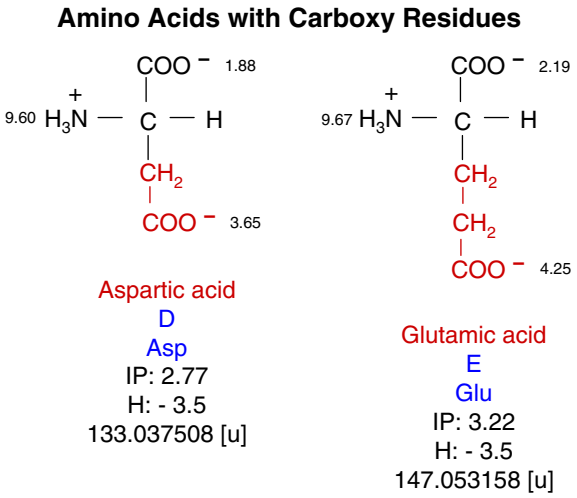

Fig. 5 – Proteinogenic amino acids with carboxy residues. The formulas present the ionization state predominating at pH 7. The hydropathy index (H) can be used to predict the tendency of an amino acid to seek a hydrophobic (values >1) or an aqueous (values <1) environment. Ionization state, IP, pK values and hydropathy index from "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

2. The chemistry of amino acids

2.1. *General aspects*

Amino acids are zwitterions because they have at least one amino group yielding a positive charge when protonated and at least one carboxylic acid group providing a negative charge if deprotonated. At physiological pH both charges are present. With respect to the functional side chain groups amino acids can have basic properties like arginine (isoelectric point (IP): 10.76; Fig. 7) or acidic like glutamic acid (IP: 3.22; Fig. 5). Amino groups are protonated if the pH is lower than their

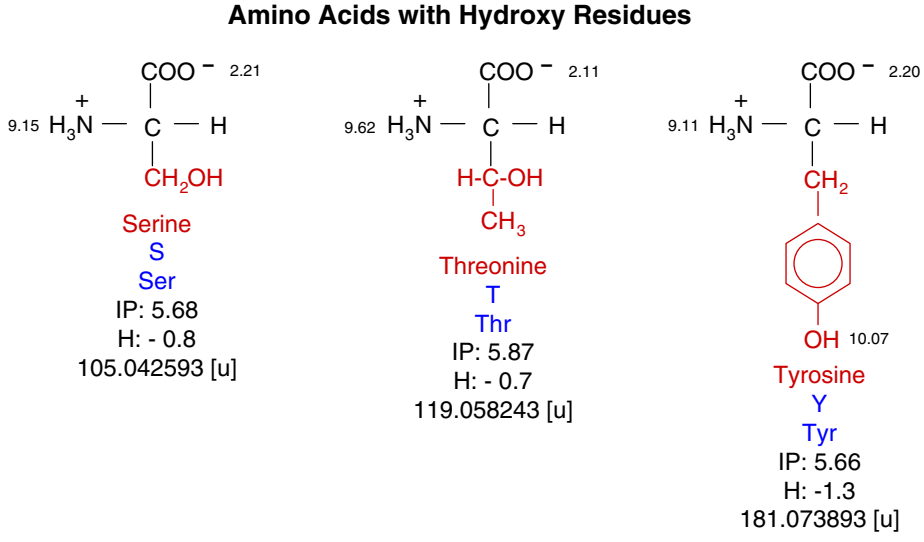

Fig. 4 – Proteinogenic amino acids with hydroxy residues. The formulas present the ionization state predominating at pH 7. The hydropathy index (H) can be used to predict the tendency of an amino acid to seek a hydrophobic (values >1) or an aqueous (values <1) environment. Ionization state, IP, pK values and hydropathy index from "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

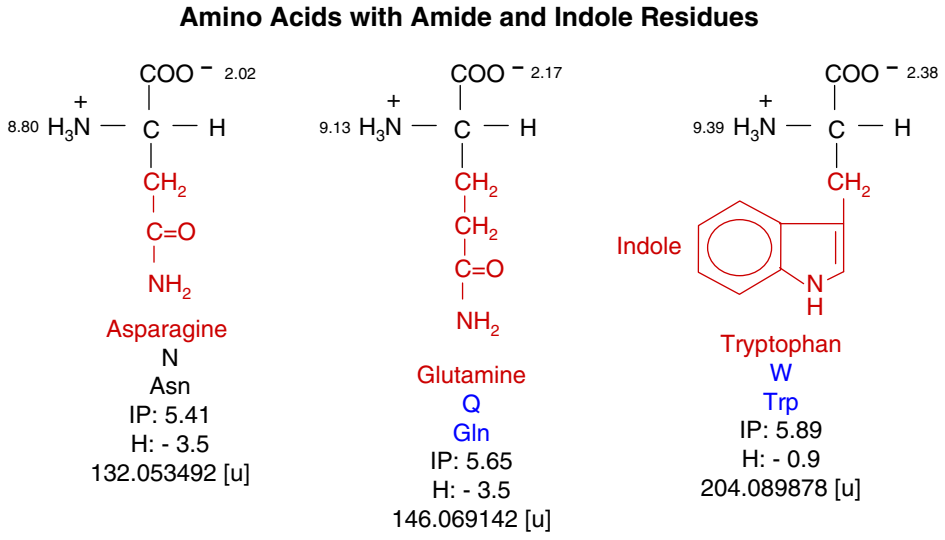

Fig. 6 – Proteinogenic amino acids with amide and indole residues. The formulas present the ionization state predominating at pH 7. The hydropathy index (H) can be used to predict the tendency of an amino acid to seek a hydrophobic (values >1) or an aqueous (values <1) environment. Ionization state, IP, pK values and hydropathy index from "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

corresponding dissociation constant (pK_a) and carboxylate groups are deprotonated and thus ionized if the pH is above the respective pK_a value. The pK_a values of the amino acid side chains are listed in Figs. 1–9. The isoelectric points critically determine the behavior of amino acids, peptides and proteins during electrophoretic or chromatographic separation and affect signal intensity in mass spectrometry. The charge state has also an effect on the chemical reactivity of the functional groups. In the unprotonated state the functional side chains of arginine, lysine, histidine, cysteine, aspartic acid, glutamic acid and tyrosine are potent nucleophiles. The relative order of nucleophilicity of functional groups in amino acids is $R\text{-}S^- > R\text{-}NH_2 > R\text{-}COO^- = R\text{-}O^-$ [7].

The most frequently occurring reaction of amino acids in living organisms is the formation of an amide bond, termed peptide bond (Fig. 10), between the α-carboxyl group of one

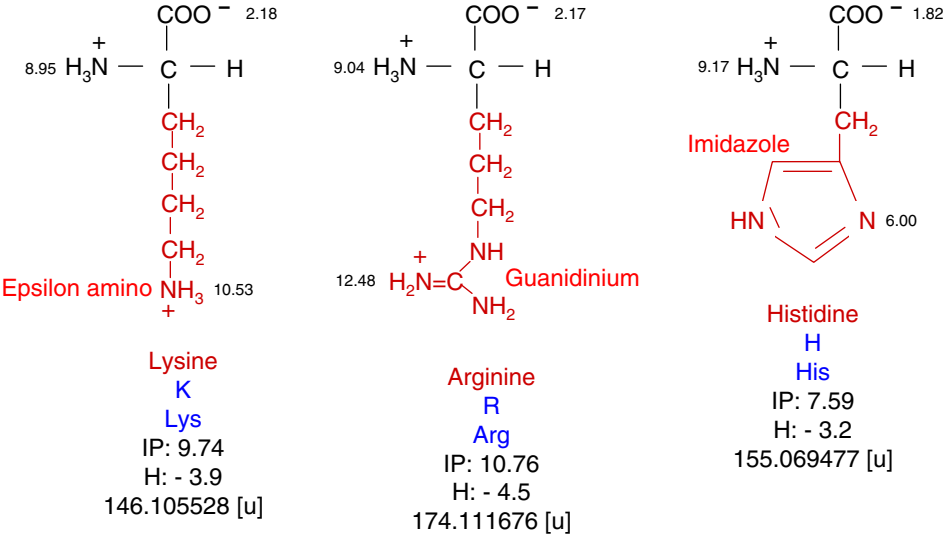

Fig. 7 – Proteinogenic amino acids with amino residues. The formulas present the ionization state predominating at pH 7. The hydropathy index (H) can be used to predict the tendency of an amino acid to seek a hydrophobic (values >1) or an aqueous (values <1) environment. Ionization state, IP, pK values and hydropathy index from "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

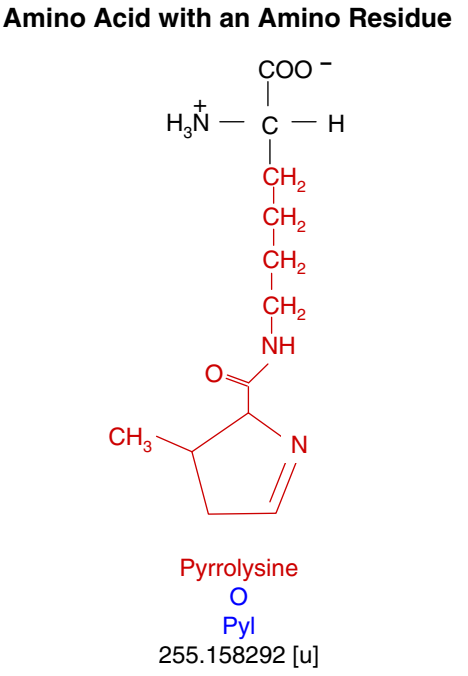

Fig. 8 – Proteinogenic amino acids with an amino residue. Pyrrolysine has not been identified in eukaryotes. The formulas present the ionization state predominating at pH 7. From "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

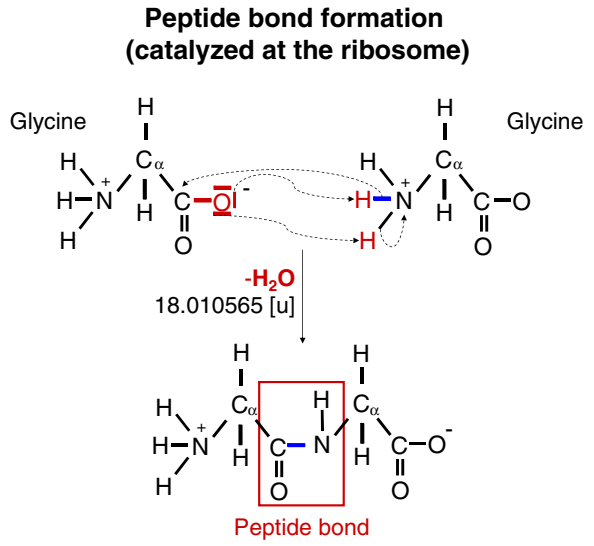

Fig. 10 – Peptide bond formation is catalyzed by the peptidyltransferase activity at the ribosome. The carbonyl group of the carboxylic acid group is activated as active ester in the form of an aminoacyl-tRNA.
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

amino acid with the α-amino group of a second amino acid or with the N-terminus of the growing peptide chain. This reaction, which is catalyzed at the ribosome by a peptidyltransferase and is repeated many times, forms the basis of protein biosynthesis. Peptide bond formation can be described as a nucleophilic attack of the α-amino group of an aminoacyl-tRNA on the activated carbon atom of the esterified carboxyl group of the peptidyl-tRNA. Redox reactions comprise a second important class of reactions of amino acid side chains. The sulfur-containing amino acids cysteine and methionine are most responsive to oxidation [7].

The general scheme of many enzyme-catalyzed reactions by which a nucleophilic side chain residue (carboxyl group, hydroxyl group, amino group, thiol group) of an amino acid in a protein is modified can be described as the covalent

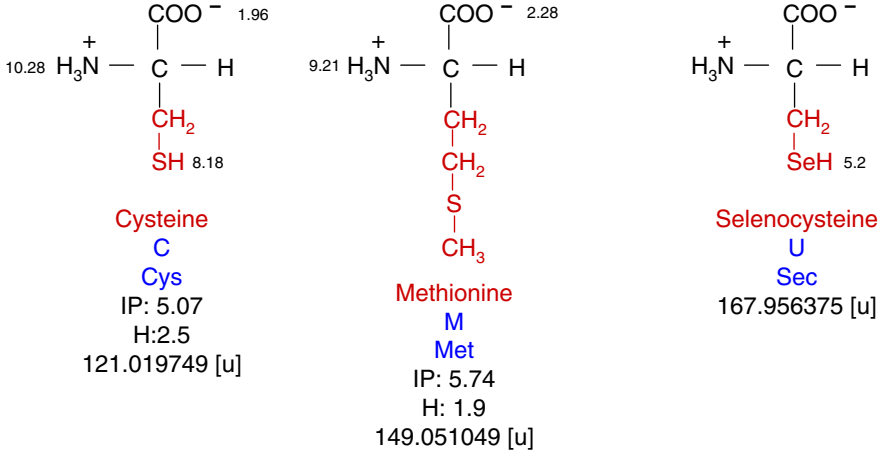

Fig. 9 – Proteinogenic amino acids with thiol, thioether or seleno residues. The formulas present the ionization state predominating at pH 7. The hydropathy index (H) can be used to predict the tendency of an amino acid to seek a hydrophobic (values >1) or an aqueous (values <1) environment. Ionization state, IP, pK values and hydropathy index from "Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York".
The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

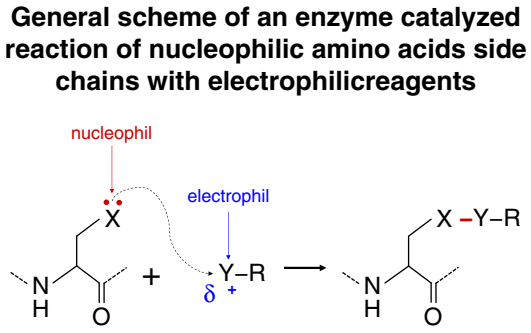

Fig. 11 – General scheme of an enzyme-catalyzed reaction by which the side chain residue of an amino acid in a protein is modified can be described as the covalent addition of an electrophilic chemical group to the nucleophilic electron-rich side chain of the amino acid.
(Reproduced with permission from C. T. Walsh et al., Angew. Chem. Int. Ed. 2005, 44, 7342–7372; copyright Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim).

addition of an electrophilic chemical group to the nucleophilic electron-rich atom in the side chain of the amino acid (Fig. 11). Usually the electrophile is part of a co-substrate, which is the donor of the electrophilic fragment. Fig. 12 presents some of the co-substrates serving as donors of electrophiles [8]. The table in Fig. 13 lists typical modifications resulting from this type of reaction and Fig. 14 gives a few examples of modified amino acids due to enzymatic reactions.

The proteinogenic amino acids in Figs. 1–9 are grouped according to the reactivity of their side chains. From the amino acids with alkyl chains (Fig. 2) very few naturally occurring derivatives are known. This is also partially true for phenylalanine (Fig. 3) having an aryl side chain. However the aryl group is more reactive than the alkyl groups; oxidation of the aryl side chain is a common reaction yielding, for example, tyrosine and L-3,4-dihydroxyphenylalanine (L-DOPA). The oxygen atom, present in the hydroxyl-group-containing amino acids serine, threonine and tyrosine (Fig. 4), is a moderate nucleophile. An important reaction of these hydroxyl groups is the formation of esters, especially with inorganic phosphate. Phosphorylation, formally the replacement of the hydrogen atom of the hydroxyl moiety with the phosphoryl group (PO_3^{2-}), requires catalysis by an enzyme (kinase) and a co-substrate in which the electrophilicity of the phosphor atom of the phosphoryl group is increased. This is the case for the γ-phosphor atom of the phosphoryl donor adenosine triphosphate when complexed with Mg^{2+} [9]. Phosphorylation of proteins is a fundamental reversible modification reaction by which the function of proteins is controlled and which plays a central role in intracellular signaling (see tutorial on Natural Post-Translational Modifications). The importance is reflected by more than 500 genes in the human genome which are proposed to be coding for kinases [10] and about 150 phosphatases responsible for removal of phosphate groups from proteins [11,12].

While the alpha-carboxylic acid and alpha-amino groups are used to form the peptide bond, a number of amino acids have reactive functional groups in their side chains that can be enzymatically or chemically modified. This leads to post-translational modifications (PTMs) with endogenous and exogenous reactants including man-made substances. Many of these PTMs are crucial to the function and regulation of proteins and most of them are reversible. There are, however, also PTMs, which occur as spontaneous chemical reactions (in the absence of enzymatic catalysis), for example, with carbohydrates, and which lead to deleterious protein modifications as in the case of advanced glycation endproducts (AGEs), which are the result of the reaction between aldehyde functionalities in monosaccharides such as glucose and amino groups in proteins. Formation of N^6-carboxymethyl lysine (Fig. 15) is an example of such a modified amino acid. AGE-modified proteins are notably found when glucose levels are not well-controlled in diabetic patients but their presence is ubiquitous due to the presence of monosaccharides and proteins, peptides or free amino acids in the same biochemical compartments. From a proteomics point of view these modifications are

Electrophilic Co-substrates

Cosubstrate (Donor substrate)	Activated electrophil	Nucleophilic atom in the amino acid	Type of modification	Class of enzymes	Enzymes removing the modification
ATP	δ+$PO_{3(phosphoryl)}$	O (Ser, Thr, Tyr, Asp) N (His)	Phosphorylation	kinases	phosphatases
Acetyl-Coenzyme A (Acetyl-CoA)	O ‖ CH_3C δ+	N-epsilon (Lys)	Acetylation	acetyltransferases	deacetylases
Myristoyl-CoA	O ‖ $CH_3(CH_2)_{11}CH_2C$ δ+	N (α-Aminogroup) S (Cys)	Myristoylation (acylation)	N-myristoyltransferase S-myristoyltransferase	myristoyl hydrolase[1]
S-Adenosylmethionine methylation	δ+CH3	N (Lys, Gln, His, Arg) O (Asp, Glu) S (Cys) C (Arg, δC; Gln, αC)	Methylation	methyltransferase	Not known
NAD^+ [1]	ADP-ribose (δ+C1-atomofthe ribose)	N (Arg, Asn) O (Asp, Glu, Ser) S (Cys)	ADP-ribosylation	ADP-ribosyltransferase	ADP-ribose-protein hydrolase

[1] not yet described
[2] ref. 8

Fig. 12 – Non-comprehensive overview over co-substrates that serve as electrophilic donors for amino acid side chain modifications. (Adapted from Rucker RB, Wold F. Cofactors in and as posttranslational protein modifications. FASEB J. 1988 Apr;2(7):2252–61).

	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Lys	Met	Pro	Ser	Thr	Trp	Tyr
Acylation				■					■						
ADP-ribosylation	■	■	■	■		■						■			
Carboxylation						■									
Disulfide formation				■											
Glycosylation		■					■					■			
Hydroxylation									■		■				
Isomerisation			■												
Mannosylation														■	
Methylation	■					■		■	■						
Nitration														[1]	■
Nitrosylation				■											
Oxidation				■						■					
Prenylation				■											
Phosphorylation			■	[2]				■				■	■		■
Sulfation															■
Transglutaminat.					■										

Fig. 13 – Common enzyme catalyzed reactions of amino acids in proteins (the table is based on reference [13]) (gray indicates that this amino acid is susceptible to this reaction); [1] ref. [30]; [2] ref. [12].

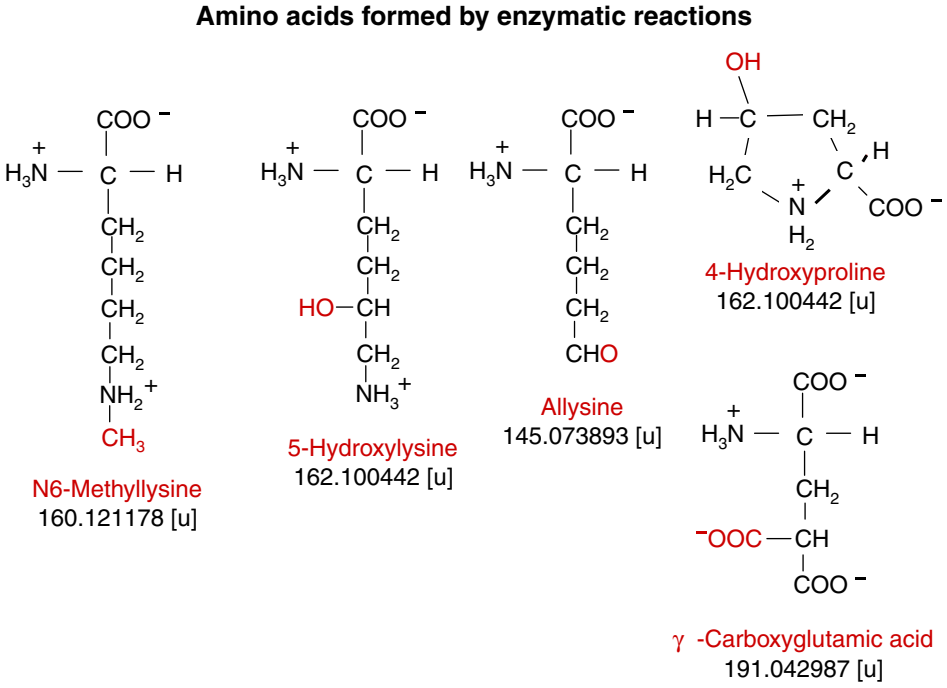

Fig. 14 – Examples of modified amino acids, which are generated from proteinogenic amino acids by the action of enzymes. N6-Methyllysine, 5-hydroxylysine and allysine are found mostly in collagen and generated from lysine by oxidation catalyzed by lysyl oxidase. Hydroxyproline is a major component of collagen. Hydroxyproline is formed by hydroxylation of proline by the enzyme prolyl hydroxylase. γ-Carboxyglutamic acid is part of a number of proteins that are involved in blood coagulation and is formed under the action of γ-glutamyl carboxylase with vitamin K as co-factor. The formulas present the ionization state predominating at pH 7.
From: Lehninger Principles of Biochemistry. 2000. 3rd ed. Nelson DL, Cox MM. Worth Publishers. New York. The exact monoisotopic mass is given in u (source: http://pubchem.ncbi.nlm.nih.gov).

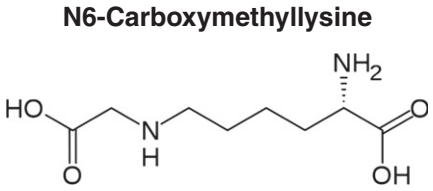

Fig. 15 – Example of an advanced glycation endproduct (AGE) resulting from the reaction of lysine with carbohydrates and subsequent oxidation by reactive oxygen species (ROS). N^6-carboxymethyllysine (CML) is a modified amino acid that is found in proteins with increasing age, in diabetic patients or related to atherosclerosis.

changing the molecular mass of proteins and peptides as well as their physical–chemical properties such as charge or hydrophobicity, which affects retention time upon chromatographic separation, ionization efficiency and detection by mass spectrometry.

PTMs, including proteolytic protein processing and protein splicing, result in a constantly changing proteome, which is much more dynamic and complex than the corresponding genome [13]. Analyses, which rely on comparison of experimental data with sequence databases, are thus not able to capture the diversity and dynamics of the proteome when it comes to PTMs. Based on the physical–chemical properties, it is however, possible to enrich proteins and peptides containing certain PTMs and the combination with mass spectrometry has led to the discovery of more than 200 different modifications be they natural or unnatural [13–15]. Recently developed high-resolution mass spectrometers and the use of different scan modes, for example in triple quadrupole instruments, and fragmentation principles open many possibilities to detect, identify and ultimately quantify PTMs even in highly complex biological samples.

Detection and identification of PTMs, notably uncommon ones, in proteins and peptides require sequencing in combination with sophisticated search algorithms that take all possible interpretations of a given set of MS/MS spectra into account [16–22]. Some functional side chains of amino acids can be labeled for affinity-based enrichment as well as for specific detection of the modified proteins or peptides.

2.2. *Reactions of the thiol group (sulfhydryl group) of cysteine*

The chemistry of cysteine provides a good example to introduce the reader to the enormous diversity of reactions of the functional side chains of amino acids and the resulting complexity of modifications in proteins. Reactions of cysteine are determined by its thiol group. Its chemical properties are

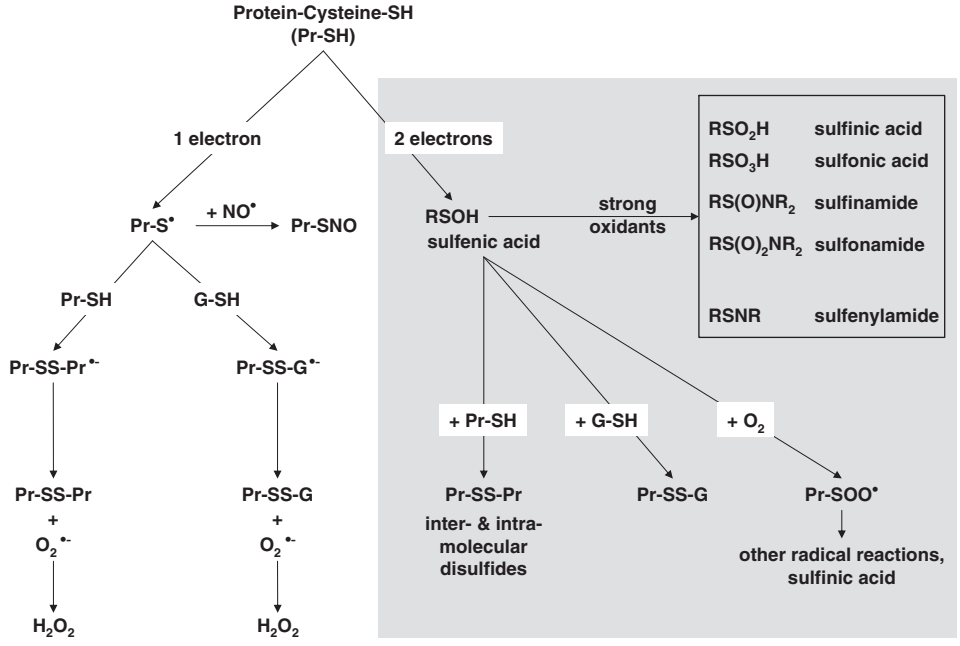

Fig. 16 – The chemistry of the thiol group of cysteine. Two electron oxidations of a protein thiol (Pr-SH), yield sulfenic acid (Pr-SOH), which is a transient intermediate. With glutathione (GSH) it forms a mixed disulfide (Pr–SS–G), with other cysteines intramolecular or intermolecular disulfides and with adjacent amides it yields a sulfenylamide by condensation. Hypohalous acids, which can be produced *in-vivo* by peroxidase-catalyzed reactions of halide ions with H_2O_2, first give rise to sulfenyl halides, which hydrolyze to the sulfenic acid. One electron oxidants, such as radicals or transition metal ions, form the thiyl radical (Pr–S$^{\bullet}$). The most preferred reaction of this radical under aerobic conditions with a thiolate anion (GSH or protein-SH) yields the disulfide anion radical, which perpetuates the reaction with oxygen forming superoxide. The thiyl radical can also transmit radical reactions or be quenched by scavengers (modified from ref. [22]).
(Reproduced with permission from Christine C. Winterbourn, Mark B. Hampton; Thiol chemistry and specificity in redox signaling. Free Radical Biology & Medicine 45 (2008) 549–561. Copyright Elsevier B.V.).

Generation of cysteine sulfenic acid after halogenation

$$HOCl + Cys\text{-}SH \rightarrow H_2O + Cys\text{-}SCl \rightarrow Cys\text{-}SOH + HCl$$

Fig. 17 – Reaction of cysteine with hypochlorous acid to yield the sulfenyl chloride that subsequently hydrolyzes to sulfenic acid (ref. [25]).

notably dependent on the sulfur atom, which is crucial to the central role of sulfur-containing biomolecules in redox reactions and to the presence of thiols in prosthetic groups (*e.g.* phosphopantetheine, lipoic acid) and enzyme cofactors (*e.g.* coenzyme-A, S-adenosylmethionine), where the sulfur atom participates directly in the catalytic reaction mechanisms. Compared to primary amines, thiols are more nucleophilic, in particular at physiological pH, where primary amines are largely protonated. The thiol moiety of cysteine reacts more rapidly than lysine, resulting in the possibility of selective modification of cysteine in the presence of lysine in proteins and peptides [23].

The thiol group reacts with nearly all physiological oxidants. The majority of oxidants prefer the thiolate anion ($-S^-$). Two principal oxidation mechanisms have been observed, one electron oxidations and two electron oxidations (Fig. 16). One electron oxidations result in thiyl radicals as first transient intermediates. In the presence of oxygen thiyl radicals react with thiolate anions in proteins (PrS^-) or in the ubiquitous glutathione (GS^-) to yield the disulfide anion radicals, $Pr–SS–Pr^-$ and $Pr–SS–G^-$. The presence of oxygen conducts the reaction toward the generation of disulfides (Pr–SS–Pr, Pr–SS–G) and superoxide anion radicals ($O_2.^-$), leading to further oxidative reactions [24]. Two-electron oxidations of a protein thiol (Pr-SH) yield sulfenic acid (Pr-SOH). This reactive intermediate forms mixed disulfides with glutathione (Pr–SS–G) or cysteine as well as intramolecular or intermolecular disulfide bonds. Sulfenic acid may further react with adjacent amides to yield a sulfenylamide by condensation. Hypohalous acids, which can be produced *in vivo* by peroxidase-catalyzed reactions of halide ions with H_2O_2 [25], for example by myeloperoxidase from neutrophils [26], first give rise to sulfenyl halides that hydrolyze to the corresponding sulfenic acid (see Fig. 17) [27].

One electron oxidants, such as radicals or transition metal ions, generate the thiyl radical (Pr–S.). The most preferred reaction of this radical under aerobic conditions with a thiolate anion (GSH or protein-SH) yields the disulfide anion radical, which perpetuates the reaction with oxygen forming superoxide anion radicals, as described above. The thiyl radical can also transfer the radical reaction to different neighboring molecules or it can be quenched by scavengers. Reaction of thiyl radicals with nitric oxide (NO) leads to S-nitrosylation (Fig. 18). S-nitrosylation is a principal effector mechanism in redox-based regulation of protein function including subcellular localization, molecular interactions, activity, and turnover of proteins, as reviewed in detail by Hess et al. [28]. About a thousand S-nitrosylated proteins have already been identified [29], thus underlining the importance of this cysteine modification in proteins. An excellent and comprehensive overview about the redox chemistry of the thiol group in living organisms is given by Winterbourn and Hampton [24].

Depending on the concentration of oxygen, NO and thiols, as well as redox states, different mixtures of products such as RSSR (R: residue other than a protein, *e.g.* glutathione), PrSSPr (Pr: protein), RS-NO, PrS-NO, and mixed disulfides of low molecular weight molecules (R_1SSR_2) and proteins (RS-SPr) are generated in biological compartments. In the case of a deficiency of reducing capacity or oxidant overproduction, higher level oxidative species of thiol modification, for instance sulfenic (PrSOH), sulfinic ($PrSO_2H$) or sulfonic acids ($PrSO_3H$), may be formed. Whereas PrSOH formation is reversible, that of $PrSO_2H$ and $PrSO_3H$ is irreversible and may result in tissue damage [30].

The electron-rich thiolate anion of cysteines cannot only be oxidized but may also serve as a target for the attack of electrophilic groups resulting in covalent addition of the electrophile to the thiolate anion. Donor substrates providing "building blocks" with increased electrophilicity are the coenzyme NAD offering the ADP-ribosyl moiety for ADP ribosylation (Fig. 19), farnesyl-diphosphate and geranyl-diphosphate, offering electrophilic alkyls for prenylation (Fig. 20), palmitoyl-CoA offering palmitoyl for palmitoylation (Fig. 21) and ubiquityl-AMP offering ubiquitin for ubiquitination of cyteines in E1-ubiquityl ligases (Fig. 22). The covalent addition of ubiquitin to the cysteines of E1-ubiquityl ligases is required for the transfer reaction of ubiquitin toward its target proteins. This is achieved by increasing the electrophilicity of the acyl group of C-terminal glycine by the covalent attachment to cysteine via a thioester. The activation principle using acyl thioesters is comparable to the activation system mediated by coenzyme A (CoA). Addition of ubiquitin to the thiol-group of the E1 ubiquityl ligase occurs via ubiquityl-AMP as an activated intermediate [13]. In the majority of investigated ubiquitinated proteins ubiquitin is bound via the ε-amino group of lysines to the target protein. However, it

S-nitrosylation of cysteine

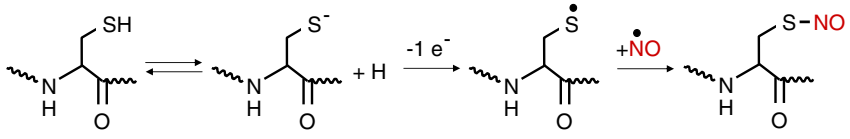

Fig. 18 – S-nitrosylation of cysteines. The thiolate radical forms S-nitrosyl-cysteines with nitrogen oxide (NO) (ref. [13]). (Reproduced with permission from C. T. Walsh et al., Angew. Chem. Int. Ed. 2005, 44, 7342–7372; copyright Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim).

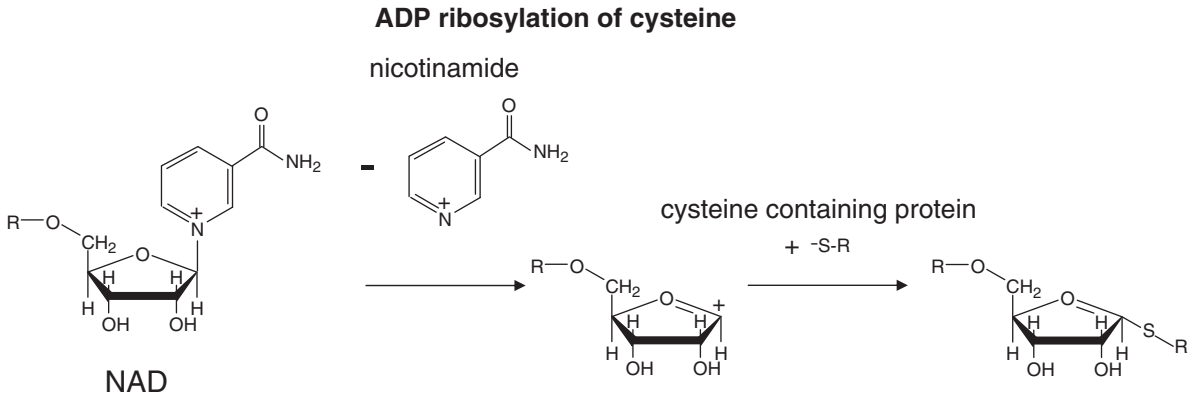

Fig. 19 – ADP ribosylation of cysteine (ref. [13]).
(Reproduced with permission from C. T. Walsh et al., Angew. Chem. Int. Ed. 2005, 44, 7342–7372; copyright Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim).

has been reported that also cysteine is used as a residue for ubiquitination [31]. Concerning the palmitoylation reaction this term does not only describe the reaction of palmitic acid with the thiolate anion but includes also other types of fatty acids (as reviewed in Levental et al. [32]). Therefore the term S-acylation is a better description for this type of reaction.

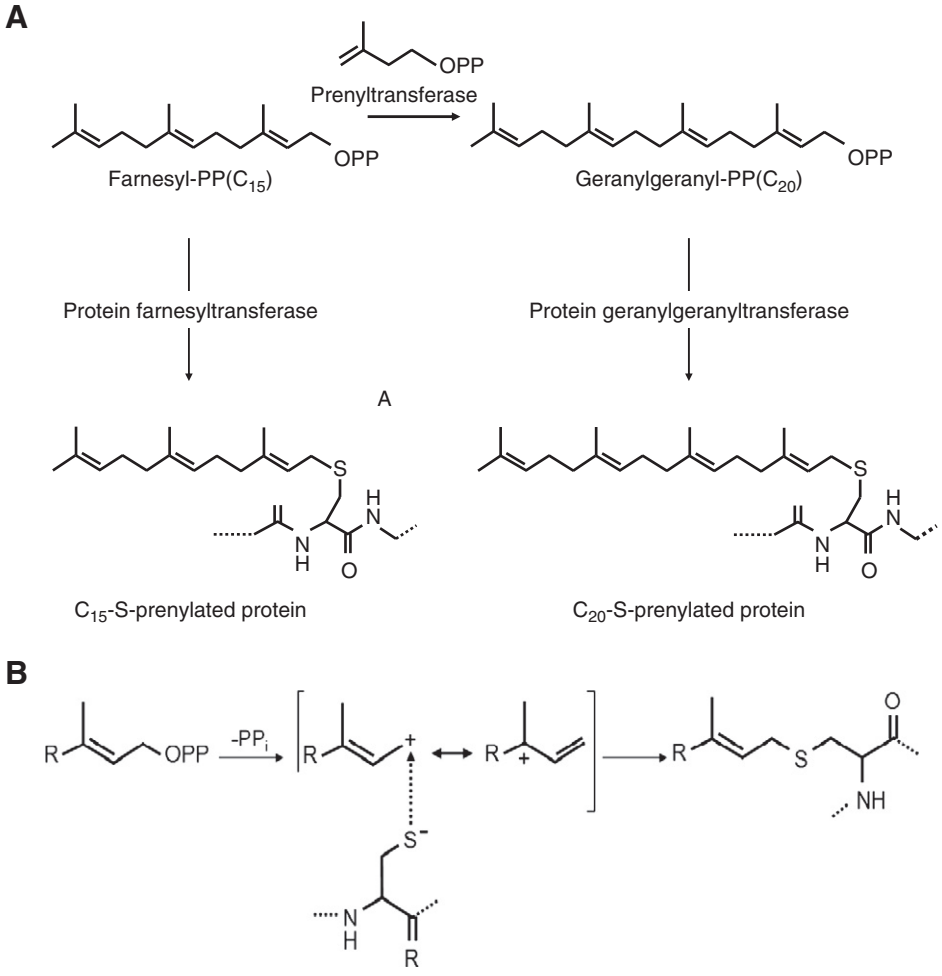

Fig. 20 – Prenylation of cysteine. This reaction type belongs to the alkylation reactions. –OPP: diphosphate moiety; PPi: inorganic diphosphate. (A) Addition of farnesyl and geranylgeranyl to cysteine. (B) Reaction mechanism of the prenylation of cysteine (ref. [13]).
(Reproduced with permission from C. T. Walsh et al., Angew. Chem. Int. Ed. 2005, 44, 7342–7372; copyright Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim).

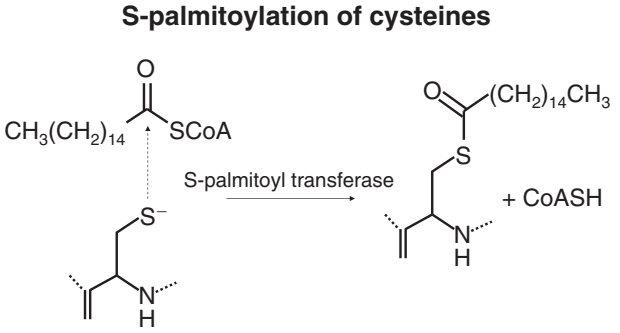

Fig. 21 – S-palmitoylation of cysteines. The electrophilicity of the acyl group of palmitate is increased due to thioester formation with coenzyme A. Note: Not only palmitic acid but also other fatty acids have been observed attached to thiol groups of cysteines (refs. [13] and [30]).
(Reproduced with permission from C. T. Walsh et al., Angew. Chem. Int. Ed. 2005, 44, 7342–7372; copyright Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim).

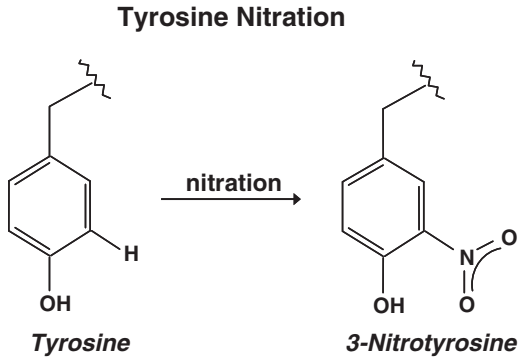

Fig. 23 – Schematic drawing showing the formation of nitrotyrosine through reaction between tyrosine and a nitrating agent.
(Reproduced with permission from Abello et al., (2009) J Proteome Res. 8, 3222–3238; copyright American Chemical Society).

2.3. *Reaction of tyrosine with reactive nitrogen species (RNS)*

Nitrotyrosine results from the reaction between the amino acid tyrosine and a nitrating agent such as peroxynitrite (Fig. 23). Similarly, tryptophan may be nitrated at the 6-position [33]. RNS are generated *in vivo* during an inflammatory reaction, for example during the oxidative burst of activated neutrophils. Nitration changes the chemical properties of tyrosine in a way that allows selective chemical labeling after reduction of the nitro group to an aromatic amine [14,34–43]. This property of nitro-tyrosine has been used to introduce a biotin moiety and to enrich biotinylated peptides (Fig. 24). Chemical derivatization of nitrotyrosine after reduction to the corresponding aromatic amine can also provide a chemical handle to introduce groups that produce characteristic fragments upon tandem mass spectrometry and MS^3 [35]. This is an example where chemical derivatization of a post-translationally-modified amino acid can serve to a) introduce an affinity handle for enrichment or b) incorporate a reporter group for selective mass spectrometric detection (Fig. 25).

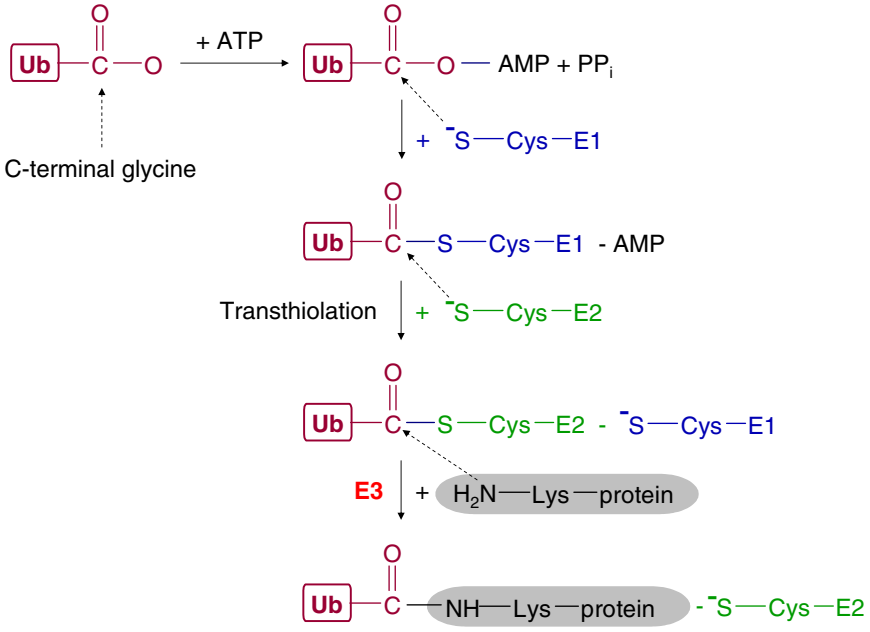

Fig. 22 – Ubiquitination. Ub: Ubiquitin. Thioester formation of acyl-ubiquitin with the cysteine residues of E1-ubiquityl ligases increases the electrophilicity of the acyl group of ubiquitin for the transfer of ubiquitin to E1-ubiquityl ligases to E2-ubiquityl ligases and from the latter to the ε-amino group of protein lysines.
(Adapted from ref. [13]).

2.4. *Deamidation of asparagine and glutamine and isoaspartic acid formation*

Glutamine and asparagine are prone to deamidation both *in vivo* and *in vitro* notably when they are followed by small amino acids such as Gly or Ser [29–37]. We focus on the deamidation of asparagine, as this is most often observed and stands exemplary for most of the reactions occurring with glutamine. Deamidation proceeds via formation of a cyclic succinimide intermediate, which subsequently opens to either L-Asp or L-Glu upon hydrolysis with water (Fig. 26). It is noteworthy that deamidation is a chemical reaction that may occur at basic as well as acidic pH values during the *in vivo* lifetime of a protein (*e.g.* for proteins that have a very slow turnover) or *in vitro* (*e.g.* during storage of proteins). It has, for example, been shown that aging of the eye lens is accompanied by modification of crystallins, among which deamidation [38].

A critical event related to deamidation is the possible ring opening of the succinimide intermediate in the "wrong" direction, meaning that the backbone of the protein continues via the side chain of aspartic acid (better isoaspartic acid) leading to major distortion of the local and possibly global protein structure, which is often accompanied by a significant loss of biological activity and function. Deamidation is not only a potential problem during *in vivo* aging of long-lived proteins

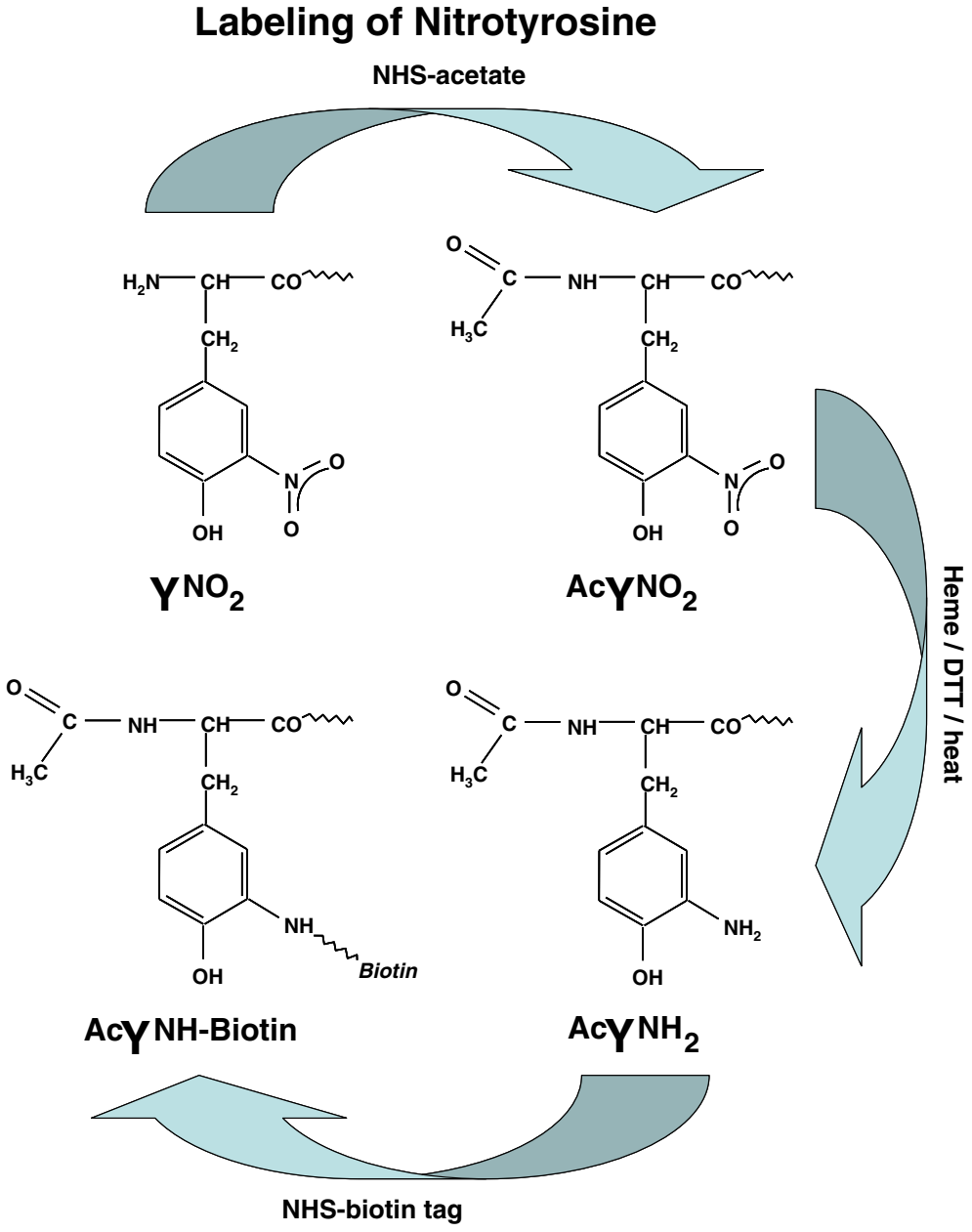

Fig. 24 – Labeling of nitrotyrosine with biotin after reduction to aminotyrosine and subsequent acylation with an activated ester of biotin after initial blockage of all free amine groups with an activated ester of acetic acid.
(Reproduced from Abello et al., (2010) Talanta 80, 1503–1512, copyright Elsevier).

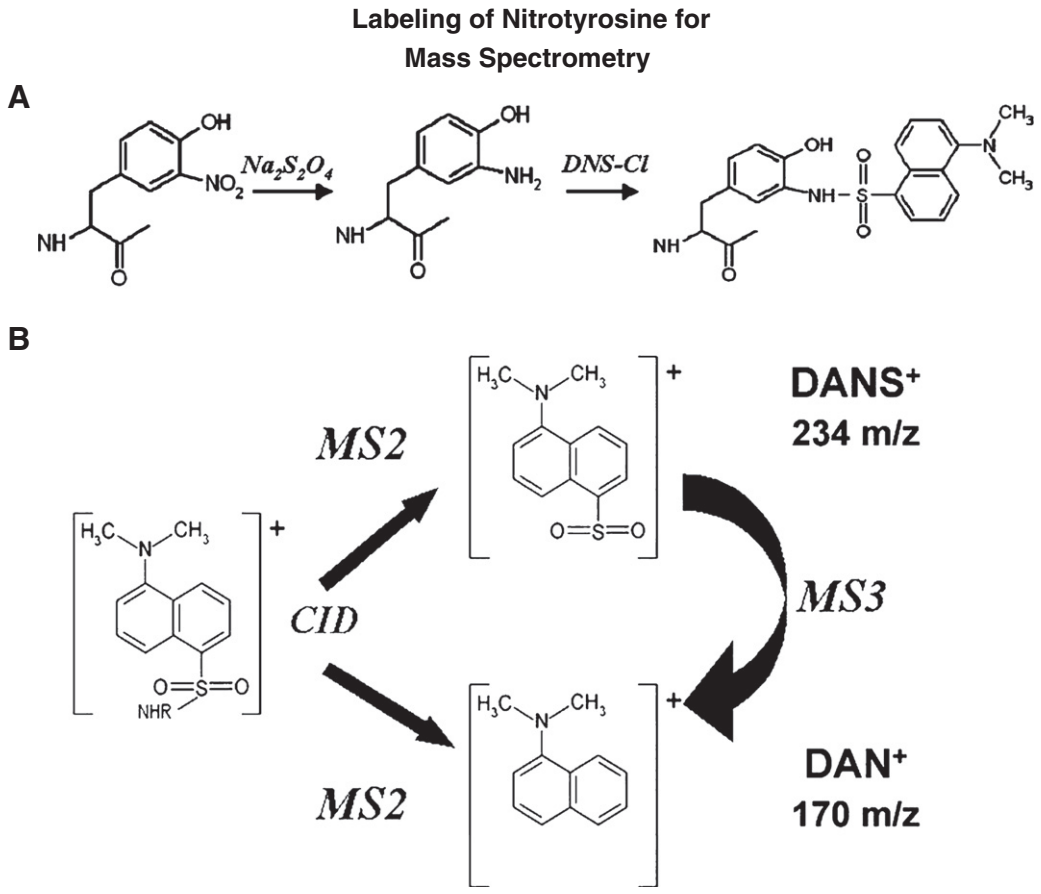

Fig. 25 – (A) Labeling of nitrotyrosine with dansylchloride after reduction to aminotyrosine after initial blockage of all free amine groups with an activated ester of acetic acid. (B) Fragmentation of labeled nitrotyrosine upon collision-induced dissociation in a quadrupole ion trap mass spectrometer to generate characteristic reporter ions. (Reproduced from Amoresano et al., (2007) Anal. Chem. 79, 2109–2117, copyright American Chemical Society).

but also in the production of proteins for therapeutic use, for example, of monoclonal antibodies. As deamidation is affected by a number of external factors such as pH and water content, it can be reduced by storing proteins as freeze-dried formulations buffered at slightly acidic pH. On the other hand, there are also inherent structural elements in proteins, such as the nature of the amino acid that is C-terminal to Asn or Gln, which play a decisive role. For qualitative and notably quantitative proteomics experiments it is therefore important to be aware of the chemical instability of Asn (and to a lesser extend Gln) and to avoid peptides containing Asn-Gly sequences for quantification, if at all possible.

The presence of isoAsp can be detected by a number of approaches (see [30,37] for reviews), notably thanks to the enzyme protein L-isoaspartyl O-methyltransferase (PIMT), which esterifies the newly formed carboxylic acid group in isoAsp with a methyl-group that is derived from S-adenosylmethionine. This reaction was originally used to radiolabel isoAsp residues in proteins allowing their sensitive detection in complex protein mixtures based on 2D gel electrophoresis followed by autoradiography. More recently mass spectrometry in combination with charge-sensitive separation techniques has become the method of choice to follow deamidation of proteins.

2.5. *Reaction of lysine with aldehydes (protein carbonylation)*

Protein carbonylation, of which AGE-modified amino acids are an example (see Fig. 15), introduces aldehyde functional groups that are generally absent in proteins, except for glycoproteins, where the aldehyde is part of the oligosaccharide chain. Protein carbonylation is related to oxidative stress, the aging of an organism or tissue and many other pathological or pre-pathological states. Comprehensive analysis of protein carbonylation may thus serve as an indicator of a developing disease, or as Regnier et al. have put it, of the biological age of an organism, organ or tissue [23–28]. Since proteins may be carbonylated at many sites, it is critical to obtain an overview by proteomic profiling in order to correlate carbonylation patterns with disease states (Fig. 27).

The fact that aldehyde groups are not part of the repertoire of functional groups in proteins makes it possible to derivatize them with chemical tags in a selective manner. Notably the hydrazide derivatives shown in Fig. 28 have contributed significantly to gaining a better understanding of the relation of protein carbonylation to various biochemical processes related to oxidative stress on a proteome-wide scale. They serve as an example of how selective derivatization techniques combined with affinity chromatography can provide a more

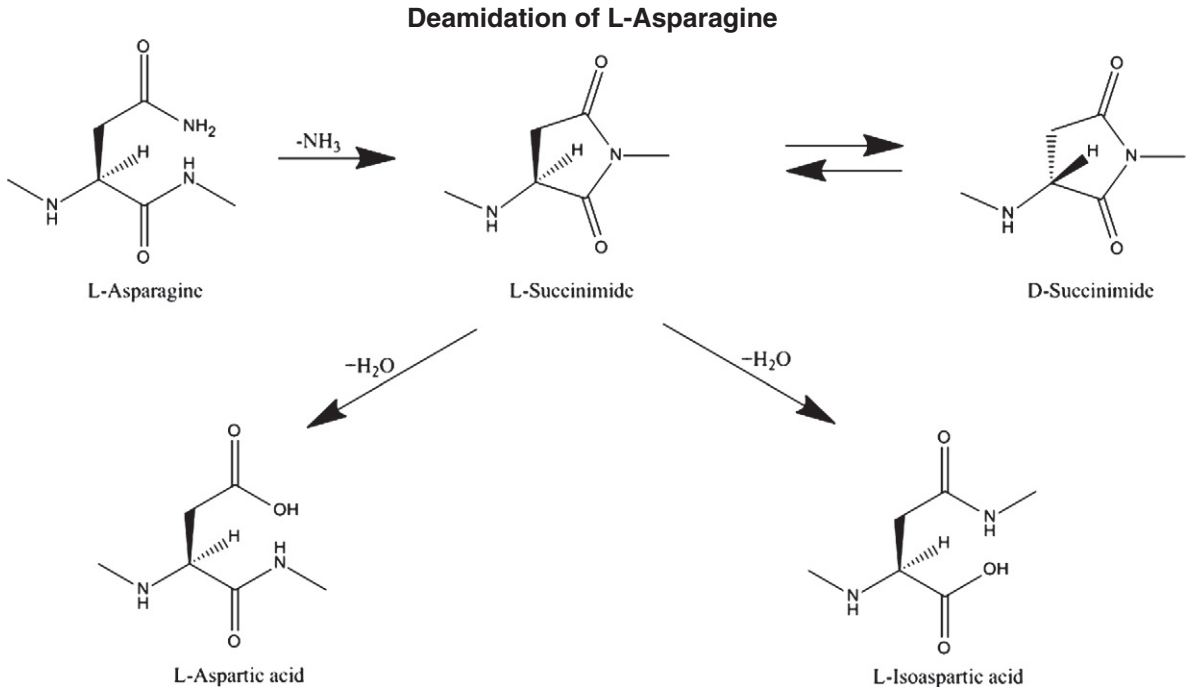

Fig. 26 – Deamidation of L-Asparagine (and glutamine) initiated through condensation of the side-chain amide group with the carbonyl group of the following amino acid (to the C-terminal side) to give a cyclic L-Succinimide. The L-Succinimide is in equilibrium with its corresponding D-form, which can undergo the same reactions as the L-form. Upon hydrolysis, the L-Succinimide ring opens to form either L-Aspartic acid or L-Isoaspartic acid. Both modifications introduce an additional negative charge in the protein and can be detected by charge-sensitive separation methods, notably isoelectric focusing gel electrophoresis or ion-exchange chromatography. Isoaspartic acid leads furthermore to a significant change in the backbone structure of the protein and is thus often accompanied by a significant or even total loss of biological activity.

detailed view of this particular amino acid modification. This example shows, however, also that unnatural PTMs represent a significant challenge to current proteomics technology and that each of them requires a dedicated approach.

3. Basic concepts to identify amino acid modifications

The number of exogenous natural and unnatural post-translational modifications exceeds 200 and it is therefore not possible to detail each and every one of them in this tutorial. We have chosen to introduce two basic concepts that have proven successful and widely applicable in identifying, localizing and quantifying unnatural PTMs in proteins based on the case of nitrotyrosine.

3.1. *Concept 1*

Tandem mass spectrometry in conjunction with liquid chromatography (LC–MS/MS) is a cornerstone technique to elucidate the structure of unnatural PTMs, to localize them in proteins and peptides and ultimately to quantify defined PTMs using different scan modes of notably triple–quadrupole mass analyzers. LC–MS/MS may be combined with chemical derivatization to enhance selectivity, sensitivity or to allow enrichment of modified proteins or peptides prior to analysis. The analysis of nitrotyrosine serves as a good example of how these techniques can be combined.

The occurrence of nitrotyrosine in proteins has received attention because of its possible involvement in aging, neurodegenerative disease or inflammatory disorders such as rheumatoid arthritis or asthma. As shown in Figs. 23 and 24, nitrotyrosine can be reduced to aminotyrosine using hemin and dithiothreitol (DTT) or sodium dithionite. To gain a better understanding of how chemical derivatization and tandem mass spectrometry can work together to provide an overview over this modification, it is interesting to focus on the derivatization of aminotyrosine with dansylchloride (Fig. 25), which results in a sulfonamide linkage between the amino group of aminotyrosine and the dansyl-tag [35]. To render this reaction selective for nitrotyrosine, it is necessary to block all other reactive amines, for example through reaction with acetic acid N-hydroxysuccinimide ester or acetic acid anhydride, prior to reduction (see Fig. 24) [37,44]. To facilitate the reaction, due to better accessibility of the nitration sites, it is advisable to perform chemical derivatization after proteolytic digestion. Once derivatized, the dansylated aminotyrosine residues must be selectively detected. Amoresano et al. achieved this through an elegant combination of precursor ion scanning with MS^3. Dansylated peptides fragment into m/z 234 and m/z 170 ions upon collision-induced dissociation (CID) in a triple quadrupole mass spectrometer (see Fig. 25B). An initial precursor ion scan searches for a defined fragment ion (in this case m/z 170) and traces this back to the precursor from which it was derived

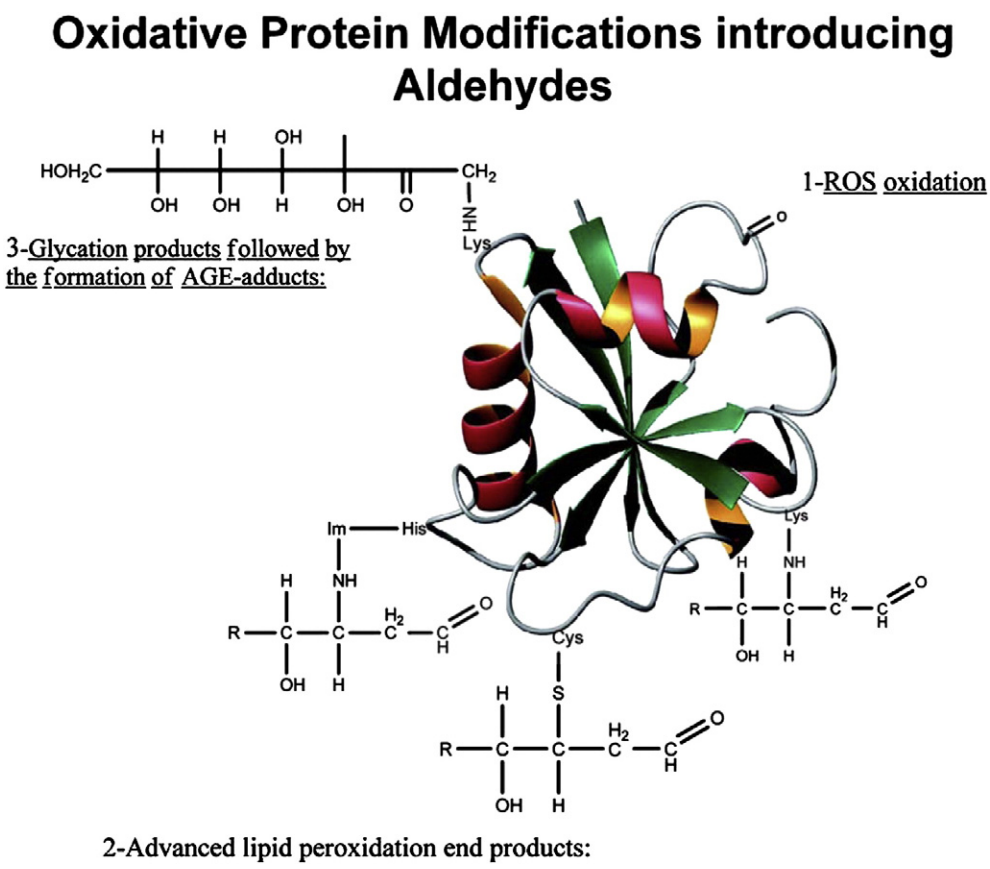

Fig. 27 – Possible oxidative modifications of a protein leading to introduction of aldehyde groups. (Reproduced from Madian and Regnier (2010) J. Proteome Res. 9, 3766–3780; copyright American Chemical Society).

(see Fig. 29 for a schematic view). Upon detection of fragment ion m/z 170, Q3 is switched to linear ion trap mode to capture the m/z 234 fragment ion and to fragment it to the m/z 170 ion in an MS^3 experiment (see Fig. 25B).

Fig. 30 illustrates the gain in selectivity of the precursor ion scan mode relative to the full-scan mode (compare traces A and B) and of the precursor ion scan mode with a subsequent MS^3 scan relative to the precursor ion scan alone (compare traces B and C). Fig. 31 shows how this approach allows detection of a number of nitrotyrosine-containing peptides in a tryptic digest of *E. coli* proteins.

This example highlights how chemical derivatization combined with optimized mass spectrometric analysis achieves a significant gain in selectivity to detect and quantify unnatural post-translational modifications in proteins. While various types of derivatization strategies alone or in combination with mass spectrometric detection routines can tackle a wide range of natural and unnatural PTMs, the approach has its limitations in terms of sensitivity and selectivity. To enhance sensitivity, it is sometimes necessary to combine LC–MS/MS with dedicated sample preparation notably based on affinity chromatography or solid-phase extraction. Selectivity of detection also benefits from sample preparation, since m/z values of precursor and/or fragment ions are not necessarily specific for a given PTM. This has been reviewed by Stevens et al. in the case of nitrotyrosine [45].

3.2. *Concept 2*

Post-translational modifications change the physical–chemical properties of proteins and peptides. This can be exploited to enrich them or separate them from the unmodified forms. Diagonal chromatography, as originally described by Cruickshank et al. [46], has been adapted and further developed by Gevaert and Vandekerckhove et al. to cover a wide range of natural and unnatural PTMs among them nitrotyrosine [15,39]. As stated above, nitrotyrosine can be reduced to amino-tyrosine. Discrimination between peptides containing aminotyrosine and unmodified peptides is possible due to a decreased retention time of the aminotyrosine-containing peptide on reversed-phase columns at slightly acidic pH due to protonation of the aromatic amino group (Fig. 32). This separation principle can be extended to entire proteome samples by first separating a tryptic digest with reversed-phase HPLC, collecting fractions, treating the peptides in each fraction with dithionite and reinjecting them on the same column under the same chromatographic conditions. All peptides that contained a nitrotyrosine will elute earlier due to conversion to aminotyrosine while the remainder of peptides will elute at the original retention time. This allows isolation and further characterization of nitrotyrosine-containing peptides as shown on the example of nitrated bovine serum albumin (Fig. 33). In case the fragment ion series covers the modified tyrosine residue, its location in the protein can be assigned.

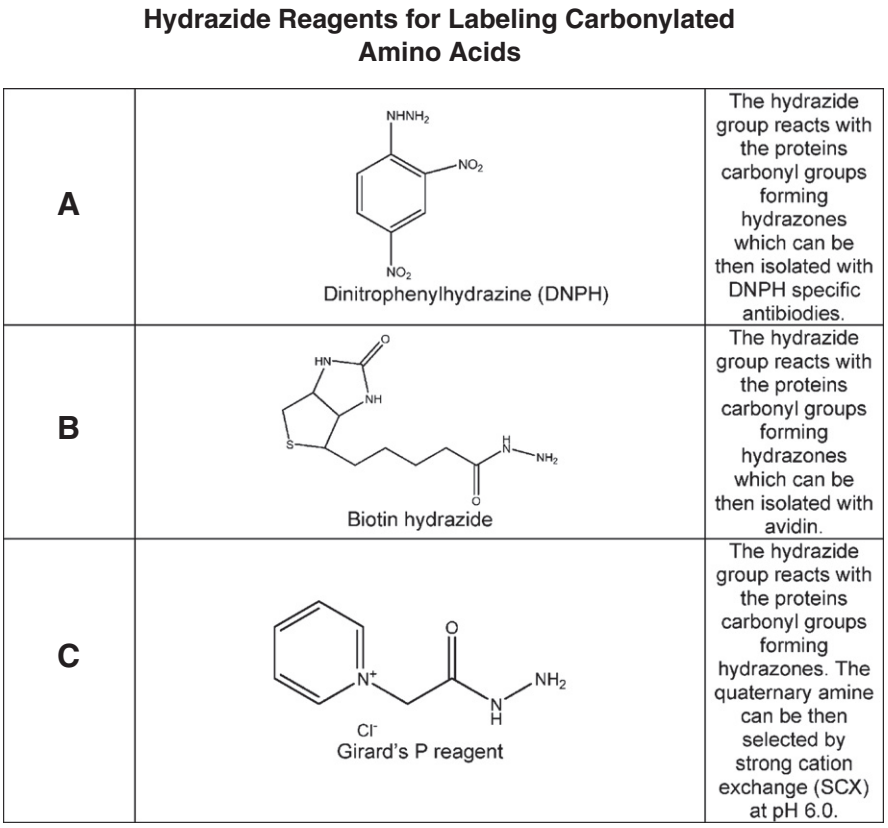

Fig. 28 – Selected labeling reagents for the enrichment of carbonylated proteins. (Adapted from Madian and Regnier (2010) J. Proteome Res. 9, 3766–3780; copyright American Chemical Society).

4. Future perspectives

The story about natural and notably non-natural post-translational modifications of amino acids in proteins and peptides is certainly not finished. It is likely that we will discover more unexpected modified amino acids in proteins due to their inherent chemical reactivity. While a few, highly relevant PTMs, such as the phosphorylation of serine, threonine and tyrosine, have attracted much attention, many others go undetected by every day, routine proteomics workflows. This is due to the way proteomics is currently

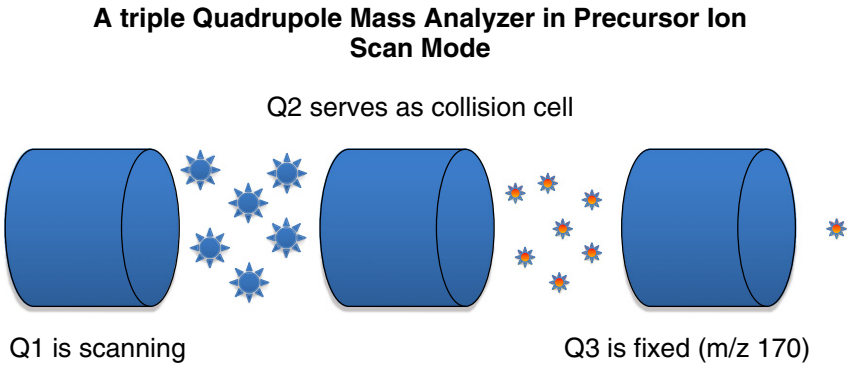

Fig. 29 – Schematic representation of a triple quadrupole mass spectrometer operating in precursor ion scan mode. Quadrupole 1 (Q1) is scanning a given, user-defined range of m/z values (precursor ions) while Q3 is fixed at the value of a common fragment ion (in the case of dansylated nitrotyrosine m/z=170; see ref. [33]). Q2 operates as a collision cell for collision-induced dissociation (CID) to generate the fragment ions. In this mode of operation, the detector will only give a response if a precursor ion generates a fragment ion of m/z=170. In this particular case, Q3 can be switched to function as a linear ion trap allowing to capture the other dansyl-specific fragment ion of m/z=234 (see Fig. 25B) and fragment it to the m/z=170 ion. By combining these scan modes, it is possible to detect dansylated peptides (or any other dansylated compounds for that matter) in a highly specific manner (see Fig. 28 for an example).

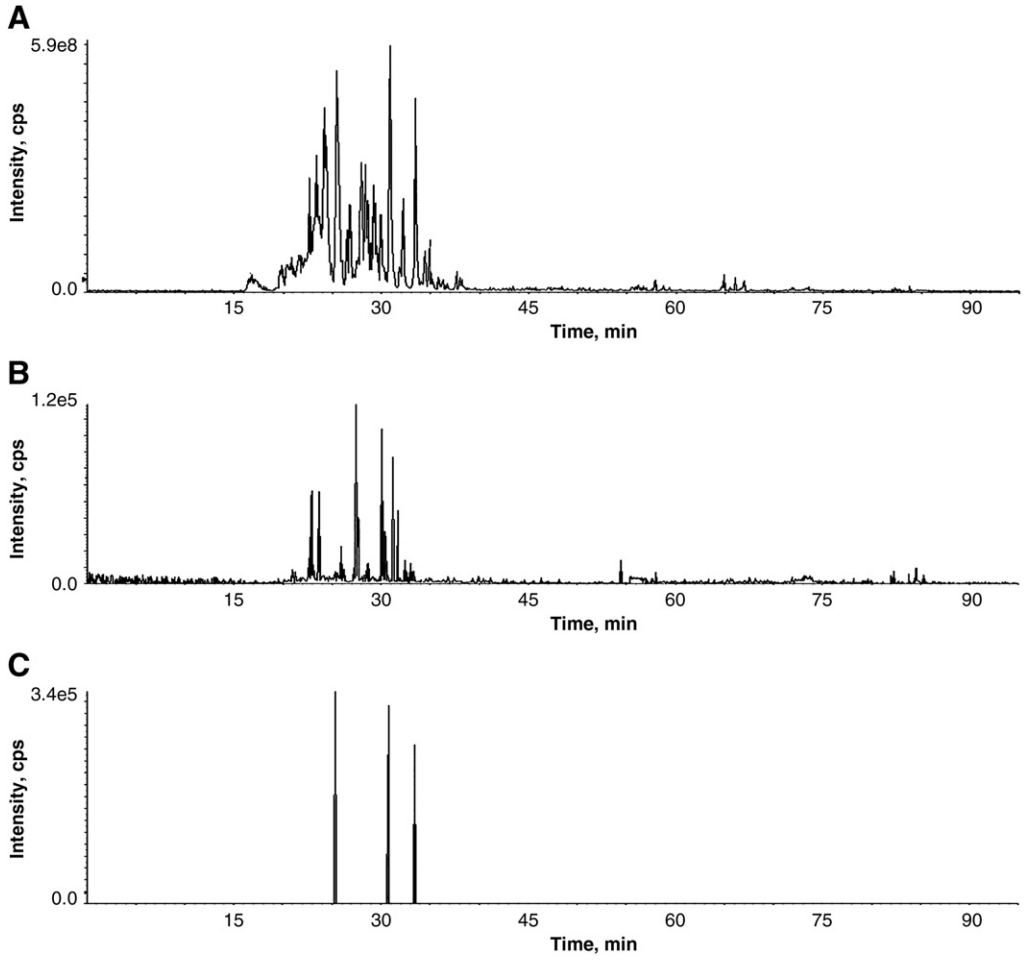

Fig. 30 – LC–MS/MS analyses of trypsin-digested, dansylated, tyrosine-nitrated bovine serum albumin. (A) MS full scan profile, (B) m/z 170 precursor ion scan profile and (C) total ion current profile for the transition m/z 234 to m/z 170 in Q3 (MS^3 mode). (Reproduced from Amoresano et al., (2007) Anal. Chem. 79, 2109–2117, copyright American Chemical Society).

performed, in that experimental data are compared with expected patterns based on information that is available in protein or DNA sequence databases. Anything that is "unexpected" in the experimental data (generally tandem mass spectra obtained by CID or more recently also by ETD) remains unassigned and thus unexplained. As most proteomics laboratories work with automated data processing and analysis workflows, due to the enormous amount of generated data, there is little manual checking of unassigned mass spectra unless there is a compelling reason to search for a particular modification. It is thus unlikely that routine proteomics will contribute significantly to the discovery of new, unexpected PTMs, since one is searching "under the lamppost".

Collaboration between experimentalists and bioinformaticians is starting to change this picture. Experimentalists and notably analytical protein chemists are devising new chemical labeling strategies to tackle PTMs and bioinformaticians work on better data analysis software that is not restricted to matching amino acid sequences in databases with the obtained experimental data but that try to interpret spectra for which there is no obvious database hit [47,48]. Mass spectrometers with higher resolution and mass accuracy and separation systems with higher resolution make important contributions in that they provide higher quality raw data.

We could only scratch the surface of the vast area of amino acid modifications in this tutorial but hope to have achieved one thing; to sensitize the reader to the diversity of the proteome, which contains still many uncharted domains with relevance to a better understanding of biology in general and of disease mechanisms in particular. We hope to stimulate more proteomics researchers to think "out of the box" and not to accept that about 50% of their good-quality MS/MS spectra do not match the entries in a given database when analyzed with standard search algorithms.

5. A systematic terminology for describing modified amino acids in proteins

The ability to identify modified amino acids in proteins calls for a terminology for the exact and unambiguous description of the analyzed proteins. A systematic terminology was first suggested by Schluter et al. [49], following the aim to provide a system, which allows a comprehensive description of the

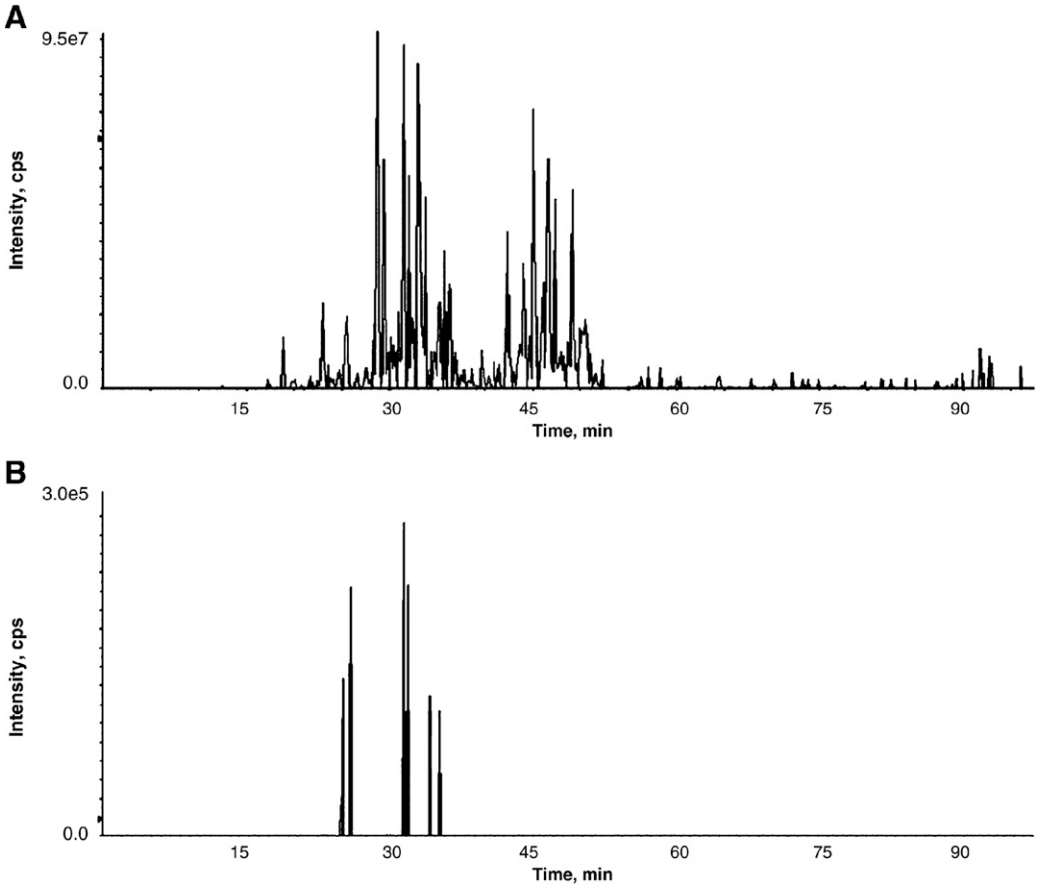

Fig. 31 – LC–MS/MS analysis of a trypsin-digested *E. coli* protein extract that was spiked with tyrosine-nitrated bovine serum albumin. (A) Reconstructed ion chromatogram of the precursor ion scan for m/z 170, (B) total ion current profile for the transition m/z 234 to m/z 170 in Q3 (MS^3 mode).
(Reproduced from Amoresano et al., (2007) Anal. Chem. 79, 2109–2117, copyright American Chemical Society).

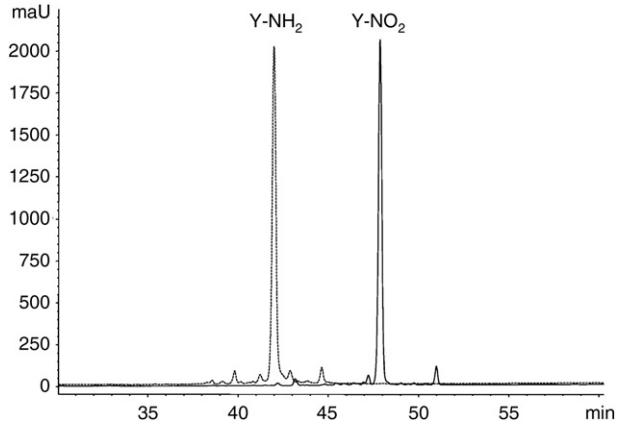

Fig. 32 – Separation of a nitrotyrosine-containing peptide and the same peptide containing aminotyrosine after reduction with sodium dithionite by reversed-phase HPLC at pH 5.5 (detection at 214 nm).
(Reproduced from Ghesquiere et al., (2009) Mol. Cell. Proteomics 8, 2642–2652, copyright the American Society for Biochemistry and Molecular Biology).

exact chemical composition of a "protein species" and which, at the same time, is as simple as possible. Application of the "protein species" terminology does not necessarily require full knowledge of the exact chemical composition of a protein species but can already be used for the description of a protein species, for which only minor experimental data about its composition are available, *e.g.* lists of some of its modified peptides. With top-down protein-sequencing, as pioneered by Kelleher and coworkers [50], the aim of identifying the exact chemical composition in total of protein species is getting more and more accessible. The term "protein species" was first suggested by Jungblut et al. in 1996 [51] and was defined for proteomics in more detail in 2008 [52]. The introduction of the term "protein species" was necessary because the term "isoform" was already defined by the nomenclature rules of IUBMB for gene products from two different genes which have the same function [53]. The protein species terminology is designed for storing the description of the exact chemical composition of a protein species in databases. This terminology enables to correlate a defined function of a protein species with its exact chemical composition. The protein species terminology is flexible and adaptable to every level of knowledge and to every level of experimental data of individual protein species. As a minimum description the entry

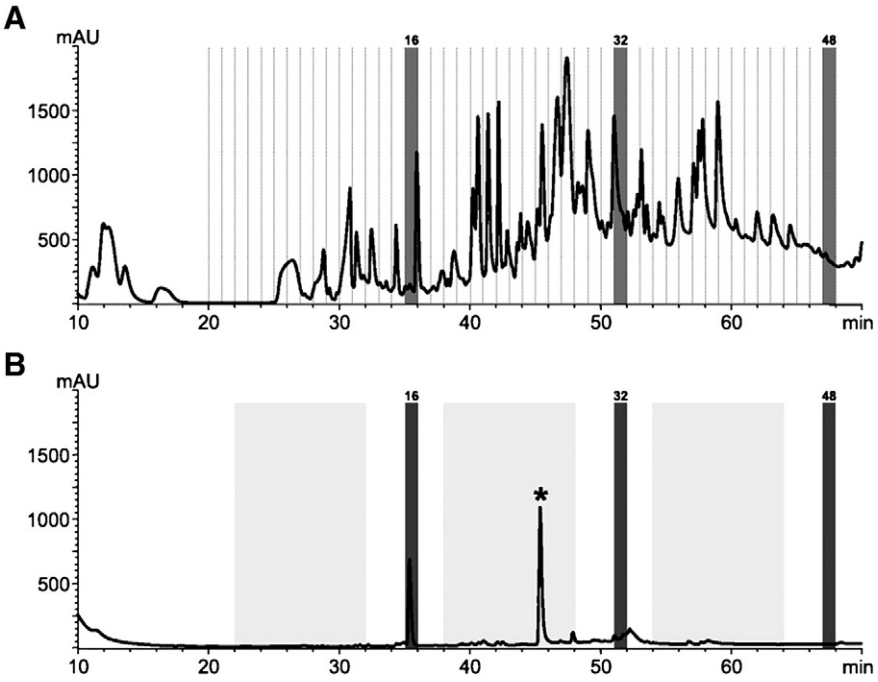

Fig. 33 – (A) Primary reversed-phase HPLC separation of trypsin-digested, nitrated bovine serum albumin (BSA). The highlighted fractions were pooled and subjected to reduction with sodium dithionite. (B) Reinjection of the pooled fractions after reduction shows that fraction 32 contained a nitrotyrosine-containing peptide, which is now distinguishable due to the retention time shift (detection at 214 nm).
(Reproduced from Ghesquiere et al., (2009) Mol. Cell. Proteomics 8, 2642–2652, copyright the American Society for Biochemistry and Molecular Biology).

name (gene name+species according to the UniProt knowledgebase) can be used, if no analytical data about the target protein species were determined. As a result the terminology reflects the depth of analysis concerning the exact chemical composition of defined protein species which have been investigated by experiments.

The basic rules of the Protein Species Identifier System (PSIS) were extended in 2009 and are available on-line (open access) [54]. Briefly, PSIS is organized at several levels (Fig. 34), which address different aspects of a given protein species such as its genetic origin. The first level names the gene coding for the protein species. Each level contains a descriptor, which consists of several terms. The first term of the gene-descriptor is G (for "gene"), the second is the gene entry name according to the UniProt protein knowledgebase and the third gives information about the species. For example, the complete descriptor of the gene for human endothelin-converting enzyme 2 is [G_ ECE2_HUMAN]. The terms are separated by an underscore.

If the protein species is coded by a gene with a polymorphism a descriptor abbreviated NP gives information about the polymorphism. For the description of the polymorphism an accession number according to dbSNP of the NCBI should be used. The descriptor AC focuses on the full length amino acid sequence of the protein species after its synthesis at the ribosome. It is described by the UniProt accession number. For the endothelin-converting enzyme 2 the accession number is O60344. Thus the complete AC-descriptor is [AC_ O60344].

If the protein of interest is a splicing product, it is described by the descriptor SD or the descriptor SI. SD (splicing, deleted) is used, if a sequence compared with the full length protein is missing, SI (splicing, inserted) is used for splicing variants, in which one or multiple amino acids were inserted. From ECE-2 a splicing variant is known (variant D), which has both a deleted sequence comprising amino acids 262–883 and an exchanged sequence (amino acids 162–261). The inserted sequence is: GFQKGTRQLLGSRTQLELVLAGASLLLAALLLGCL VALGVQYHRDPSHSTCLTEACIRVAGKILESLDRGVSPCEDFYQFS CGGWIRRNPLPDGRSRWNT. Therefore the following descriptors results: [SD_262-883] and [SI_161_GFQKGTRQLLGSRTQ LELVLAGASLLLAALLLGCLVALGVQYHRDPSHSTCLTEACIRVAG KILESLDRGVSPCEDFYQFSCGGWIRRNPLPDGRSRWNT]. In many cases long descriptors can be avoided, because UniProt accession numbers already exist for splicing variants. Here, the splicing variant has the accession number O60344-4. Thus, the descriptor SD or SI is not needed in this case. The splicing variant is fully described by [AC_ O60344-4].

If a protein species is identified which is a truncated form of a full-length protein like thrombin in its active form, it can be described by the descriptor T (truncated). Human activated thrombin (without any modified amino acid side chains) thus has the descriptors [G_F2_HUMAN]+[AC_ P00734]+[T_1-327]+[DB_UniProt_165].

Posttranslational modifications of amino acid side chains within a protein species are described by the descriptor P. The second term of this descriptor lists the numbers of the amino acids which are modified and the third term informs about the

Protein Species Identifier System (PSIS): Descriptors & Terms

Descriptor-Level	1st Term[1] Symbol	2nd Term: Name or description Example	3rd Term: Further description Example	Complete descriptor of the individual level
Gene	G	Gene Name [2] ECE2	Species human	[G_ ECE2_HUMAN]
Nucleotide polymorphisms	NP	Accession number[3] rs35875049	-	[NP_rs35875049]
Initial amino acid sequence [4]	AC	Accession number[5] O60344	-	[AC_ O60344]
Splicing variant, deleted sequence	SD	Number of the first and the last amino acid within the sequence which is deleted by splicing in comparison to full length sequence. 262-883 (see UniProt accession number: O60344-4)	-	[SD_262-883]
Splicing variant, inserted sequence	SI	Number of amino acid, which precedes the sequence, which was inserted by splicing 161	Sequence of the inserted peptide GFQKGTRQLLGSRTQLELVLAGASLLLAALLLGCL VALGVQYHRDPSHSTCLTEACIRVAGKILESLDRG VSPCEDFYQFSCGGWIRRNPLPDGRSRWNT (please note that this protein species can be described by the UniProt accession number O60344-4 thus making it unnecessary to list this descriptor)	[SI_161_GFQKGTRQLLGSRTQL ELVLAGASLLLAALLLGCLVALGV QYHRDPSHSTCLTEACIRVAGKI LESLDRGVSPCEDFYQFSCGGW IRRNPLPDGRSRWNT]
Truncated amino acids	T	Sequence described by the first & the last number of the amino acids within the removed sequence 1-327 (activated thrombin; F2_HUMAN]	-	[T_1-327]
Post-translational modifications	P	Amino acid(s), which are modified 39	Accession number of the post-translation modification[6] 21	[P_39_21]
Cofactors	C	Amino acid(s), which bind the cofactor 720, 724, 780	Symbol describing the cofactor Zn	[C_720-724-780_Zn]
Function	F	EC number[7] EC 3.4.24.71	-	[F_ EC 3.4.24.71]
Localization	L	Cellular localization membrane	-	[L_membrane]
Databases	DB	Name of the data base UniProt, dbSNP. UniMod, BRENDA	Version number(s)	[DB_UniProt, dbSNP, UniMod, BRENDA_107, 132, 2006.10.16, 2011.1]

The rules for PSIS are available at (open access): http://www.beilstein-institut.de/ESCEC2009/Proceedings/Schlueter/Schlueter.pdf

Fig. 34 – Tabular overview of the descriptors and terms of the Protein Species Identifier System (PSIS). The rules for the PSIS are available at http://www.beilstein-institut.de/ESCEC2009/Proceedings/Schlueter/Schlueter.pdf (open access). The terminology is based on ref. [47].

[1] Defined aspect of the chemical composition of the protein species
[2] Recommended database: UniProt
[3] Recommended database: dbSNP (NCBI)
[4] of the protein synthesized at the ribosome
[5] Recommended database: UniProt
[6] Recommended database: UniMod
[7] Recommended database: BRENDA.

type of modification. The type of modification is given as a defined number which is listed in the database UniMod. For example, Rush et al. detected a phosphorylated peptide from ECE-2 [55]. For the description of their result the combination of the following terms is sufficient: [G_ ECE2_HUMAN]+[AC_ O60344]+[P_39_21]. The descriptor P contains the information that the amino acid at position 39 is phosphorylated.

The presence of a cofactor in a protein species is indicated by the descriptor C. The metalloprotease ECE-2 contains Zn^{2+} in its active site. The complete descriptor is [C_720-724-780_Zn]. The descriptor F describes the function of a protein species. In the case of enzymes the EC number is very helpful. ECE2 is a peptidase generating endothelin from big endothelin [56]. Therefore the complete descriptor is [F_ EC 3.4.24.71]. The localization of a protein species can be described by the descriptor L. ECE2 bound to the membrane is described by [L_membrane]. The descriptor DB gives information about the databases and their versions used for the description of the terms. In the examples describing ECE2, the databases UniProt, dbSNP, UniMod and BRENDA were used. Thus the descriptor is [DB_UniProt, dbSNP, UniMod, BRENDA_107, 132, 2006.10.16, 2011.1].

Many bottom-up based proteomic studies focusing on a defined PTM, like phospho-proteomics, typically report lists of modified peptides. If only experimental data about these modified peptides were collected PSIS can also be applied, by starting with the gene-descriptor, followed by descriptor AC (full length amino acid sequence, see above) and the descriptor P (posttranslational modification, as described above). In this case, no data about a splicing variant or truncation may be available, thus their descriptors can be omitted.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.jprot.2012.01.041.

REFERENCES

[1] Harper's illustrated biochemistry. 28 ed. McGraw-Hill Professional; 2009.
[2] Vauquelin LN, Robiquet PJ. The discovery of a new plant principle in *Asparagus sativus*. Annales de Chimie 1806;57:88–93.
[3] Braconnot HM. Sur la conversion des matières animales en nouvelles substances par le moyen de l'acide sulfurique. Ann de Chim Phys Ser 1820;2(13):113–25.
[4] McCoy RH, Meyer CE, Rose WC. Feeding experiments with mixtures of highly purified amino acids. J Biol Chem 1935;112: 283–302.
[5] Zinoni F, Birkmann A, Stadtman TC, Bock A. Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. Proc Natl Acad Sci U S A 1986;83:4650–4.
[6] Srinivasan G, James CM, Krzycki JA. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science 2002;296:1459–62.
[7] Hermanson GT. Bioconjugate techniques. 2nd ed. London: Academic Press; 2008.
[8] Hassa PO, Haenni SS, Elser M, Hottiger MO. Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev 2006;70:789–829.
[9] Walsh CT. Posttranslational modifications of proteins: expanding nature's inventory. Roberts and Company Publishers; 2006.
[10] Braconi QS, Orchard S. The annotation of both human and mouse kinomes in UniProtKB/Swiss-Prot: one small step in manual annotation, one giant leap for full comprehension of genomes. Mol Cell Proteomics 2008;7:1409–19.
[11] Jackson MD, Denu JM. Molecular reactions of protein phosphatases—insights from structure and chemistry. Chem Rev 2001;101:2313–40.
[12] Guan KL, Dixon JE. Evidence for protein–tyrosine–phosphatase catalysis proceeding via a cysteine–phosphate intermediate. J Biol Chem 1991;266:17026–30.
[13] Walsh CT, Garneau-Tsodikova S, Gatto Jr GJ. Protein post-translational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed Engl 2005;44:7342–72.
[14] Abello N, Kerstjens HA, Postma DS, Bischoff R. Protein tyrosine nitration: selectivity, physicochemical and biological consequences, denitration and proteomics methods for the identification of tyrosine-nitrated proteins. J Proteome Res 2009;8:3222–38.
[15] Gevaert K, Van DP, Ghesquiere B, Vandekerckhove J. Protein processing and other modifications analyzed by diagonal peptide chromatography. Biochim Biophys Acta 2006;1764: 1801–10.
[16] Quandt A, Masselot A, Hernandez P, Hernandez C, Maffioletti S, Appel RD, et al. SwissPIT: an workflow-based platform for analyzing tandem-MS spectra using the grid. Proteomics 2009;9:2648–55.
[17] Quandt A, Hernandez P, Masselot A, Hernandez C, Maffioletti S, Pautasso C, et al. SwissPIT: a novel approach for pipelined analysis of mass spectrometry data. Bioinformatics 2008;24: 1416–7.
[18] Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open mass spectrometry search algorithm. J Proteome Res 2004;3:958–64.
[19] Hernandez P, Gras R, Frey J, Appel RD. Popitam: towards new heuristic strategies to improve protein identification from tandem mass spectrometry data. Proteomics 2003;3:870–8.
[20] Frank A, Pevzner P. PepNovo: de novo peptide sequencing via probabilistic network modeling. Anal Chem 2005;77:964–73.
[21] Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, et al. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal Chem 2005;77:4626–39.
[22] Frank AM, Savitski MM, Nielsen ML, Zubarev RA, Pevzner PA. De novo peptide sequencing and identification with precision mass spectrometry. J Proteome Res 2007;6:114–23.
[23] Basle E, Joubert N, Pucheault M. Protein chemical modification on endogenous amino acids. Chem Biol 2010;17:213–27.
[24] Winterbourn CC, Hampton MB. Thiol chemistry and specificity in redox signaling. Free Radic Biol Med 2008;45:549–61.
[25] Hawkins CL. The role of hypothiocyanous acid (HOSCN) in biological systems. Free Radic Res 2009;43:1147–58.
[26] Hampton MB, Kettle AJ, Winterbourn CC. Inside the neutrophil phagosome: oxidants, myeloperoxidase, and bacterial killing. Blood 1998;92:3007–17.
[27] Kettenhofen NJ, Wood MJ. Formation, reactivity, and detection of protein sulfenic acids. Chem Res Toxicol 2010;23:1633–46.
[28] Hess DT, Matsumoto A, Kim SO, Marshall HE, Stamler JS. Protein S-nitrosylation: purview and parameters. Nat Rev Mol Cell Biol 2005;6:150–66.
[29] Seth D, Stamler JS. The SNO-proteome: causation and classifications. Curr Opin Chem Biol 2011;15:129–36.
[30] Di SP, Franconi F, Frosali S, Di GD. Thiolation and nitrosation of cysteines in biological fluids and cells. Amino Acids 2003;25:323–39.
[31] Cadwell K, Coscoy L. Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. Science 2005;309:127–30.
[32] Levental I, Grzybek M, Simons K. Greasing their way: lipid modifications determine protein association with membrane rafts. Biochemistry 2010;49:6305–16.
[33] Bregere C, Rebrin I, Sohal RS. Detection and characterization of *in vivo* nitration and oxidation of tryptophan residues in proteins. Methods Enzymol 2008;441(339–49):339–49.
[34] Abello N, Barroso B, Kerstjens HA, Postma DS, Bischoff R. Chemical labeling and enrichment of nitrotyrosine-containing peptides. Talanta 2010;80:1503–12.
[35] Amoresano A, Chiappetta G, Pucci P, D'Ischia M, Marino G. Bidimensional tandem mass spectrometry for selective identification of nitration sites in proteins. Anal Chem 2007;79:2109–17.
[36] Amoresano A, Chiappetta G, Pucci P, Marino G. A rapid and selective mass spectrometric method for the identification of nitrated proteins. Methods Mol Biol (Clifton, N J) 2008;477: 15–29.
[37] Chiappetta G, Corbo C, Palmese A, Galli F, Piroddi M, Marino G, et al. Quantitative identification of protein nitration sites. Proteomics 2009;9:1524–37.
[38] Ghesquiere B, Goethals M, Van Damme J, Staes A, Timmerman E, Vandekerckhove J, et al. Improved tandem mass spectrometric characterization of 3-nitrotyrosine sites in peptides. Rapid Commun Mass Spectrom 2006;20:2885–93.
[39] Ghesquiere B, Colaert N, Helsens K, Dejager L, Vanhaute C, Verleysen K, et al. *In vitro* and *in vivo* protein-bound tyrosine nitration characterized by diagonal chromatography. Mol Cell Proteomics 2009;8:2642–52.
[40] Nikov G, Bhat V, Wishnok JS, Tannenbaum SR. Analysis of nitrated proteins by nitrotyrosine-specific affinity probes and mass spectrometry. Anal Biochem 2003;320:214–22.
[41] Radabaugh MR, Nemirovskiy OV, Misko TP, Aggarwal P, Mathews WR. Immunoaffinity liquid chromatography-tandem mass spectrometry detection of nitrotyrosine in biological

fluids: development of a clinically translatable biomarker. Anal Biochem 2008;380:68–76.
[42] van Haandel L, Killmer J, Li X, Neich C, Stobaugh JF. Phenylisothiocyanate as a multiple chemical dimension reagent for the relative quantitation of protein nitrotyrosine68 ed. ; 2008. p. 507–16.
[43] Zhan X, Desiderio DM. Nitroproteins from a human pituitary adenoma tissue discovered with a nitrotyrosine affinity column and tandem mass spectrometry. Anal Biochem 2006;354:279–89.
[44] Abello N, Kerstjens HA, Postma DS, Bischoff R. Selective acylation of primary amines in peptides and proteins. J Proteome Res 2007;6:4770–6.
[45] Stevens Jr SM, Prokai-Tatrai K, Prokai L. Factors that contribute to the misidentification of tyrosine nitration by shotgun proteomics. Mol Cell Proteomics 2008;7:2442–51.
[46] Cruickshank WH, Malchy BL, Kaplan H. Diagonal chromatography for the selective purification of tyrosyl peptides. Can J Biochem 1974;52:1013–7.
[47] Ahrne E, Muller M, Lisacek F. Unrestricted identification of modified proteins using MS/MS. Proteomics 2010;10:671–86.
[48] Ahrné E, Nikitin F, Lisacek F, Müller M. QuickMod: a tool for open modification spectrum library searches. J Proteome Res 2011;10:2913–21.
[49] Schluter H, Apweiler R, Holzhutter HG, Jungblut PR. Finding one's way in proteomics: a protein species nomenclature. Chem Cent J 2009;3(11):11.
[50] Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SMM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature 2011;480:254–8.
[51] Jungblut P, Thiede B, Zimny-Arndt U, Muller EC, Scheler C, Wittmann-Liebold B, et al. Resolution power of two-dimensional electrophoresis and identification of proteins from gels. Electrophoresis 1996;17:839–47.
[52] Jungblut PR, Holzhutter HG, Apweiler R, Schluter H. The speciation of the proteome. Chem Cent J 2008;2(16):16.
[53] Liébecq C. Joint Commission on Biochemical Nomenclature IUPAC–IUBMB: nomenclature of multiple forms of enzymes. 2nd ed. Colchester: Portland Press; 1992.
[54] Schluter H, Holzhutter HG, Apweiler R, Jungbluth PR. Suggestions for a Protein Species Identifier System. In: Hicks MG, Kettner C, editors. Experimental standard conditions of enzyme characterizations; 2009. p. 79–86.
[55] Rush J, Moritz A, Lee KA, Guo A, Goss VL, Spek EJ, et al. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol 2005;23:94–101.
[56] Emoto N, Yanagisawa M. Endothelin-converting enzyme-2 is a membrane-bound, phosphoramidon-sensitive metalloprotease with acidic pH optimum. J Biol Chem 1995;270:15262–8.